Individual Counterfactuals with Multidimensional Unobserved Heterogeneity*

Richard Blundell[†] Dennis Kristensen[‡] Rosa Matzkin[§]

December 2020

Abstract

We develop a new method for identifying and estimating counterfactuals in nonparametric models with nonseparable multidimensional unobserved heterogeneity. The method can be used when the value of a vector of interdependent variables depends in an unspecified way on vectors of observable regressors and unobservable variables. For changes in the values of the observable regressors, we identify the new values of the interdependent variables when values of the unobserved variables stay fixed. No functional form restrictions are imposed on the response function, other than invertibility on the vector of unobserved variables. The regressors can be either discrete or continuously distributed. Identification is constructive, leading to an estimator that is easily computed. The estimator possesses an asymptotically normal distribution. We apply the method using UK Kantar homescan data to estimate the heterogeneous responses to changes in prices and total expenditure of households that were observed making different choices on one budget. We find significant differences in their responses.

JEL: C20, D12

Keywords: simultaneous equations, nonseparable models, constructive identification, nonparametric methods, consumer behaviour, structural demand functions.

*Earlier drafts of this paper circulated under the title "Consumer Demand with Unobserved Heterogeneity". We would like to thank the following for helpful comments and suggestions: Xiaohong Chen, Ian Crawford, Alfred Galichon, Jinyong Hahn, Jim Heckman, Stefan Hoderlein, Arthur Lewbel, and Whitney Newey, among others, as well as seminar participants at many seminars and conferences, including Harvard/MIT, 2015 Conference on Consumer Demand at BU, Cemmap/UCL, University of Bristol, University of Chicago, California Econometric Conference, Second International Conference of the Society for Economic Measurement, 2015 World Congress of the Econometric Society, PSU, UBC, SFU, NYU, Columbia, UPF, and 2017 Nemmers Conference. The research is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. Funding from the ESRC (grant number ES/T014334/1) is gratefully acknowleged. Blundell acknowledges support from the ERC under the MicroConLab grant. Kristensen acknowledges support from the European Research Council (grant no. ERC-2012-StG 312474). Matzkin acknowledges support from NSF (grants BCS-0852261, SES-1062090, and SES-0833058). Material from the Kantar homescan data are made available by Kantar Worldpanel and used by permission at IFS. We thank Kantar for access to the data and Martin O'Connell for generous help in assembling the data for this study. The usual disclaimer applies.

[†]Department of Economics, UCL, and Institute for Fiscal Studies. E-mail: r.blundell@ucl.ac.uk

[‡]Department of Economics, UCL and Institute for Fiscal Studies. E-mail: d.kristensen@ucl.ac.uk.

[§]Department of Economics, UCLA. E-mail: matzkin@econ.ucla.edu.

1 Introduction

Suppose that we observe an individual making a choice on an observed choice set and we would like to know what that specific individual would have chosen if faced with a different choice set. It is reasonable to assume that the individual has unobserved tastes that determine his/her choices and that these unobserved tastes are heterogenous across individuals. Suppose we have such data on individuals but we do now know how the unobserved tastes enter into the decision of individuals. In this paper we develop a method that uses such data to obtain an estimate for the individual's choice under a different choice set when his/her unobserved tastes stay fixed.

As an example, we can consider different price regimes on product categories together with samples of households for which we observe some of their characteristics together with how much they spent on the product categories. Based on this sample, our method allows to predict the demand response of one or more of those households to a counterfactual increase in the prices of one or more product categories, without making any restrictions on the households' demand function other than an invertibility assumption.

The method we develop can be used not only for studying counterfactuals as the specific ones described above, but more generally for estimating nonseparable models with multidimensional unobservables and with no functional restrictions on the function of interest other than invertibility. Since these models can be interpreted as reduced form functions generated from systems of simultaneous equations, the methods contribute as well to the literature on structural estimation of nonseparable simultaneous equation models, by providing a method that does not require restrictions that have been imposed in previous works.

Interest in nonseparable models has been receiving increasing interest. These models allow much more flexibility in the way unobservable variables enter into the model. They also allow one to benefit from restrictions of economic theory that might be lost with additional specifications. For example, in a consumer demand model, economic theory implies that the demand function of any individual satisfies the Slutsky conditions. These conditions are not necessarily satisfied by demand models where unobservables enter additively unless strong additional conditions are imposed (McElroy (1987), Brown and Walker (1989), Lewbel (2001)).

When the nonseparable unobservable is a scalar, there exist methods, as the ones in Matzkin (2003), for estimation of individual counterfactuals leaving the value of the unobservable fixed, without imposing functional restrictions. Assuming that unobserved heterogeneity is a scalar is rarely appropriate, however, for models where responses are interrelated. In the consumption example, an individual unobserved taste for fresh foods will have an effect on the individual's expenditure share on prepared foods. This effect would be separate from the effect that the individual's unobserved taste for prepared foods may have on such expenditure. In models of demand and supply, the observed quantities and prices, which are determined by equilibrium conditions, depend on unobservable characteristics of both the demand and the supply equations as well as possibly on unobserved market conditions affecting both. Employing an approach where the unobserved variables are aggregated into one for each equation would impose unrealistic conditions that could generate inconsistent estimators for counterfactuals. Existent nonparametric methods that allow for multidimensional unobservables require restrictions and specifications that may not be satisfied. The objective of this paper is to propose a nonparametric method that avoids such restrictions and allows for multidimensional unobservables.

Our identification results are constructive. They lead to easily computed, natural estimators of the unknown functions of interest, which impose no restrictions even when the regressors are discrete. We develop the new nonparametric estimators and analyze their asymptotic properties. These are in turn used to carry out inference on individual responses to counterfactual changes in the values of regressors, when the values of the nonseparable unobservable variables stays fixed. We also present methods to improve the practical performance of the estimators in a number of relevant scenarios, such as when the number of observable individual characteristics is large.

Our method is based on two main assumptions, invertibility of the response function in the unobservables and a restriction on the conditional density of the vector of unobserved variables, given observable external variables that do not enter into the response function. In the consumer example, the external variables could be observable characteristics of the individual consumer. In a demand and supply example, the external variables may be characteristics of the consumers and firms or of the market. Our assumptions allow us to define a mapping between the derivative of the conditional density of the *observable* variables and the derivative of the conditional density of the *unobservable* variables, where the derivatives are with respect to the external variables. Only when responses correspond to the true value of the unobservable variables, both these derivatives equal zero. Finding such zeros in the distribution of observable variables allow us to trace responses corresponding to any fixed value of the vector of unobservable variables. The process also provides a way of testing the assumed restrictions on the conditional density of the vector of unobserved variables. External variables have been used, of course, in other models as well. For example, Berry, Levinsohn and Pakes (2004) model unobserved individual heterogeneity as functions of observed and unobserved consumer characteristics, while Cunha, Heckman and Schennach (2010) show how proxies can be used in identification of latent factors.

We apply our method to estimate the counterfactual expenditure responses to changes in prices and total expenditure of three households observed to have chosen three different expenditure shares when faced with a common vector of prices and total expenditure. We use the Kantar homescan consumer panel data from the UK, which in addition to expenditures on different food items includes an extensive set of observed household attributes, behaviours and attitudes. We use the latter to construct the external variables we use in our method. We find that the responses of the three individuals are significantly different. By comparing the results to those obtained using the same data to estimate an Almost Ideal demand model, we further demonstrate the benefit of our method.

The remainder of the paper is as follows. In the next section, we compare our method to existent related literature. Section 3 presents the main assumptions and identification results. Section 4 discusses how to extend the results to cases where the observed regressors are endogenous, to cases where interest lies on the structural equations in simultaneous equations models, and to cases where only partial identification is possible. Section 5 applies the general theory to consumer demand as considered in the empirical application. Section 6 develops the estimators and their asymptotic properties. Section 7 provides several results that are useful for implementing in practice the new methods. Section 8 presents the empirical application, and Section 9 concludes.

2 Relation to the literature

The methods developed in this paper can be interpreted as providing a nonparametric multivariate version of univariate nonseparable models, such as those developed in Matzkin (2003). In the basic univariate model, the value of a dependent variable is determined by a function whose arguments are a vector of regressors and a nonadditive unobservable variable. The function is strictly monotone in the unobservable variable but it is otherwise unknown. The unobservable variable is distributed independently, or conditionally so, of the vector of regressors, with an unknown continuous distribution. An example would be one where the observable dependent variable corresponds to a consumer's expenditure on a single good, the regressors are the price of the good and the income of the consumer, and the unobservable variable represents the individual's taste for the good. Under the strict monotonicity and independence assumptions above, the change in the individual's expenditure of the good as a response to a change in the values of the price and the consumer's income, when the individual's unobserved taste does not change, is identified. With additional restrictions, the individual's taste itself is identified. One such additional restriction fixes the conditional distribution of the unobservable to be Uniform on $(0, 1)$. In such case, the nonparametric function is the conditional quantile function of the dependent variable given the regressors, and the change in the value of the function when the value of the regressors change while the value of the unobservable stays fixed is the quantile treatment effect of Lehman (1974), see Imbens and Newey (2009). Applications of these methods include estimation of the demand for food in Blundell, Kristensen and Matzkin (2014), estimation of the demand for gasoline in Blundell, Horowitz and Parey (2017), and estimation of individual consumer surplus by Hoderlein and Vanhems (2017). The nonseparable approach with scalar unobserved heterogeneity has been extended in several directions including triangular models with endogenous regressors (Chesher, 2003, Imbens and Newey, 2009), Torgovitsky (2015), D'Haultfoeuille and Février (2015)), models with fixed effects (Altonji and Matzkin (2005)), and nonlinear difference-in-difference type models (Athey and Imbens (2006)).

In this paper, we are concerned with a multivariate version of the above model where instead of a scalar dependent variable, one is interested in a vector of interdependent variables. In such case, any unobserved variable that may affect one of the dependent variables will potentially affect the values of all of them. We are then interested in a vector function such that each of its coordinate functions depends on the whole vector of regressors and a whole vector of unobservable variables. Unlike nonparametric nonseparable triangular models, which can be identified using a sequence of univariate models each depending on only one additional unobserved variable, in our model all unobservables affect all dependent variables. Only under strong restrictions our model is observationally equivalent to either unidimensional or triangular models (Blundell and Matzkin (2014)).

Extending the univariate conditional quantile method, Carlier, Chernozhukov, and Galichon (2016) and Chernozhukov, Galichon, Henry, and Pass (2020) developed identification and estimation results for nonseparable invertible models where a vector of dependent variables is determined by an unknown vector function of observable and unobservable variables. Their methods are based on extensions of results used in optimal transport theory (Brenier (1991), McCann (1995), Villani (2003)) and follow results by Ekeland, Galichon and Henry (2012) and Galichon and Henry (2012) on multivariate quantiles. Their results require fully specifying the distribution of the vector of unobservables, and imposing the restriction that the vector function is the gradient of a convex function. The method we develop in this paper does not require these conditions. In particular, we do not assume that the covariances between the unobservable variables are known. Neither do we impose symmetry of the matrix of cross-partial derivatives, which gradients of differentiable functions do.

Our method is based on a transformation of variables equation, as used in Matzkin (2008) for developing identification results for systems of simultaneous equations. The transformation of variables equation allows establishing identification at particular values of the function of interest even when such function is not identified at some other values or uniqueness of solutions to moment conditions is not guaranteed. Identification results for simultaneous equations following this approach have been discussed and/or developed, among others, in Matzkin (2007, 2008, 2013, 2015) and Berry and Haile (2014, 2016, 2018)[1]. Of these, only Matzkin (2015) developed estimation methods, and none of these papers apply their methods in an empirical study. All the constructive identification results presented in these papers require an additive structure between unobservables and functions of observable variables, which are distributed independently of the unobservables and are individually assigned to each coordinate of the unobservables. Identification of counterfactuals using the system of reduced form functions generated from these systems of simultaneous equations would require first imposing such restrictions. The method that we develop in this paper for identifying counterfactuals avoids such restrictions. Moreover, unlike Matzkin (2015), our method can be used to estimate counterfactuals due to discrete changes in regressors, when the values of the unobservables stay fixed.

Besides developing a new method for identification of the values of the dependent variables when the values of the observable conditioning variables change while the values of the unobservables stay fixed, we also provide new results for identification of derivatives of structural functions under such changes. The new results avoid the additive structures and statistical independence assumed in the previous constructive methods that are based on a transformation of variables equation, and can be used when the regressors are discrete. The methods also avoid completeness assumptions, which are usually required to identify structural functions in systems with simultaneity using conditional model conditions. The latter include Ai and Chen (2003), Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), and Darolles, Fan, Florens, and Renault (2011) for models with additive unobservable terms, and Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2007), Horowitz and Lee (2007), Chen and Pouzo (2012), and Chen, Chernozhukov, Lee, and Newey (2014) for models with nonadditive unobservable terms.

Models with nonadditive unobservables have been considered since at least Hurwicz (1950), and unobserved heterogeneity were key elements in Heckman (1974), McFadden (1974), Heckman and Willis (1977), Lancaster (1979), and others that followed. A commonly used specification with nonadditive unobservables is a linear random coefficient model (Hildreth and Huock (1968), Swamy (1970)). The literature extending these models is very large. They include Beran and Hall (1992), Beran and Millar (1994), Beran, Feuerverger and Hall (1996), Feuerverger and Vardi (2000), Foster and Hahn (2000), and Hoderlein, Klemelae, and Mammen (2010)) for linear models, Masten

[1] Berry and Haile (2014) also develop results using the moment based method in Newey and Powell (2003).

(2018) and Hoderlein, Holzmann, and Meister (2017) for simultaneous equations, Hausman and Wise (1978), Ichimura and Thompson (1998), Bajari, Fox, Kim, and Ryan (2012), and Gautier and Kitamura (2013)) for discrete choice models, Berry, Levinsohn and Pakes (2004), Berry and Haile (2009), Chiappori, Komunjer, Kristensen (2009), and Fox and Gandhi (2011)) for differentiated products models, and many other papers and models. Nonparametric extensions of random coefficient structures include Matzkin (2003, Appendix A), Briesch, Chintagunta, and Matzkin (2007), Hoderlein, Nesheim, and Simoni (2016), and Lewbel and Pendakur (2017). All these methods impose structures that our method does not impose. Our method can, however, incorporate extensions allowing features of the models in those papers.

Our application to consumer demand requires invertibility of demand on unobserved variables, which has been studied in several papers. Berry, Gandhi and Haile (2013) provided conditions directly on demand function. We provide conditions on the utility functions, based on Brown and Matzkin (1998), Beckert (2006), and Beckert and Blundell (2008).

An alternative approach to study nonparametric heterogeneous consumer demand, first studied by McFadden and Richter (1991) and McFadden (2005), is based on restrictions on the distribution of demand generated from heterogeneous consumers. (See Matzkin (2007), and Hoderlein and Stoye (2015)) for specific cases.) Kitamura and Stoye (2018) developed a method based on these restrictions, which does not require a particular structure on the way unobserved tastes enter into preferences. Their method assumes a distribution of demand on a finite set of budgets sets is given, and uses revealed preference conditions to partition the set of budgets into subsets that are rationalizable. Unlike our method, it cannot be used to identify the demand of a specific consumer and to exploit revealed preference conditions on any given individual.

3 Framework and main results

We present in this section our model, assumptions and main results. We let $Y = (Y_1, ..., Y_{d_Y})'$ be a vector of response variables of an individual satisfying

$$Y = m(X, \varepsilon), \tag{1}$$

where $X = (X_1, ..., X_{d_X})'$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_{d_\varepsilon})'$ are vectors of, respectively, observable and unobservable variables. The function $m(x, e)$ is unknown to us and so is the distribution of the unobserved random variable ε. We impose no parametric restrictions on these two objects.

The observed individual could, for example, be a consumer with Y containing the quantities consumed of a set of different goods, X the prices of the goods and the individual's income, and ε

representing the consumer's unobserved tastes/preferences. In this case $m(X,\varepsilon)$ is the individual's demand function. Given data of (Y,X), we are then interested in identifying counterfactual demand responses.

In the first scenario under consideration, we observe a particular individual choosing $Y=y$ when $X=x$. We then wish to identify the counterfactual response in Y to a change in X while the value of the individual's ε stays constant. That is, with e being the unknown value of the given individual's ε and x' being the counterfactual value of X, the current response satisfies $y=m(x,e)$ and we wish to identify the "new" response $y'=m(x',e)$. In the demand example, this corresponds to identifying the change in demand of a given individual in response to price and/or income changes, while keeping the individual's tastes unchanged.

In the second scenario, we wish to identify the full mapping $(x,e)\mapsto m(x,e)$ together with the distribution of ε. Similarly to the univariate case, this will require stronger conditions compared to the first scenario.

With $\mathcal{X}$ denoting the support of X, our first assumption guarantees that the unobserved value of ε corresponding to an observed choice y of Y when $X=x$ is unique:

Assumption 1 For any given $x\in\mathcal{X}$, the function $e\mapsto m(x,e)$ is thrice continuously differentiable with inverse $r(y,x)$. That is,

$$Y=m(X,\varepsilon)\Leftrightarrow\varepsilon=r(Y,X).$$

Assumption 1, or variations of it, are commonly met in the literature on nonseparable models with either univariate or multidimensional unobservables. Note that Assumption 1 implicitly restricts the unobserved variables, ε, to be of the same dimension as Y, $d_Y=d_\varepsilon$. (Our model can be incorporated into one with a larger dimension of unobservables, such as in Matzkin (2003, Appendix A), Matzkin (2012) or Lewbel and Pendakur (2017).) Our identification results stated in this section only require $m(x,e)$ to be differentiable, but when analyzing the properties of the estimators developed later we will require it to be thrice differentiable. For simplicity, we maintain the stronger smoothness condition throughout.

The counterfactual under the first scenario described above takes a particular form under Assumption 1: Recall that, in terms of the function m, the counterfactual is the difference between the two choices, $m(x',e)-m(x,e)$. We denote this by $\widetilde{\Delta}_y(x,x')$. Since $y=m(x,e)$ and, by Assumption 1, $e=r(y,x)$,

$$\widetilde{\Delta}_y\left(x,x'\right):=m\left(x',r(y,x)\right)-y. \tag{2}$$

Under the second scenario, we wish to identify the counterfactual changes as indexed by $\varepsilon = e$ taking the form

$$\Delta_e(x, x') = m(x', e) - m(x, e).$$

Importantly, the computation of this last quantity require us to identify the unobserved value e of ε for the given individual of interest. In contrast, $\widetilde{\Delta}_y(x, x')$ does not involve this component and so can be identified under weaker restrictions.

If m and r were known, one could calculate $\widetilde{\Delta}_y(x, x')$ and $\Delta_e(x, x')$. However, these functions are unknown. In the univariate case, where $d_Y = d_\varepsilon = 1$, Matzkin (2003) showed that if ε is distributed independently of X with a strictly increasing distribution, the counterfactual $\widetilde{\Delta}_y(x, x')$ is identified without additional restrictions either on the function r or on the distribution of ε. This is in contrast to the case where the focus is on the identification of the *function* $m(x, e)$, and in particular on identifying how this value varies as a function of e, in which case additional restrictions are needed.

Moving beyond the scalar case, so that Y and ε are multivariate, is far from trivial. Even when the distribution of ε is fully specified, the function m and the counterfactual $\widetilde{\Delta}_y(x, x')$ may not be identified (Benkard and Berry (2006))[2]. One of our aims in this paper is to provide a pointwise method to identify $\widetilde{\Delta}_y(x, x')$ without imposing either functional restrictions on the function $m(x, \varepsilon)$, other than invertibility, or specifying the distribution of ε. To achieve this goal, we make use of additional d_Z observed variables, that we collect in a vector Z with support $\mathcal{Z} \subseteq \mathbb{R}^{d_Z}$. The variables in Z are external in the sense that they do not enter the model explicitly as given in eq. (1). At the same time, we will assume that Z is related to ε, so that we can employ observations on Z to obtain the "most likely" value of ε corresponding to a value of Z. Given a specific value z of Z, we find such value of ε by solving the "score function,"

$$\frac{\partial \log f_{\varepsilon|Z}(e|z)}{\partial z} = 0 \quad \text{or equivalently} \quad \frac{\partial f_{\varepsilon|Z}(e|z)}{\partial z} = 0 \tag{3}$$

where we have assumed that the density of ε conditional on Z exists and is differentiable with respect to z.

In the consumer demand example, where ε represents a vector of tastes, Z may be a vector of observable socioeconomic characteristics. Then, the value e satisfying the equation for a particular vector of socioeconomic characteristics z is the vector of tastes that is interpreted as the most likely for consumers with such vector of socioeconomic characteristics. Variables that are usually used

[2]Consider the example in Benkard and Berry (2006), where $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ possesses a standard Normal distribution $N(0, I)$ and where the first and second rows of a 2×2 matrix $A(x)$ are, respectively, $(\cos(x), \sin(x))$ and $(-\sin(x), \cos(x))$. The models $Y = \varepsilon$ and $Y = A(x)\varepsilon$ generate identical conditional distributions of Y given X but very different counterfactuals when the value of x varies.

as either proxies or measurements of unobservable variables are good candidates for Z; our score condition is consistent with either of these interpretations. A more detailed discussion of the role of Z in relationship to ε is provided in Section 4.4.

Equation (3) can be considered as implicitly defining one or more values of e, for given z, or implicitly defining one or more values of z, for given e. To consider the possibilities of multiple solutions, we introduce the associated solution mappings,

$$\Lambda\left(z\right) := \left\{e \in \mathcal{E} : \frac{\partial f_{\varepsilon|Z}\left(e|z\right)}{\partial z} = 0\right\}, \tag{4}$$

and

$$\Lambda^{*}\left(e\right) := \left\{z \in \mathcal{Z} : \frac{\partial f_{\varepsilon|Z}\left(e|z\right)}{\partial z} = 0\right\}. \tag{5}$$

For our first result in this section, we will assume that, at least for the value e of ε satisfying $y = m\left(x, e\right)$ for specific observed values y of Y and x of X, both $\Lambda^{*}\left(e\right)$ is a singleton and, for $z = \Lambda^{*}\left(e\right)$, $\Lambda\left(z\right)$ is a singleton. In other words, we will first assume that both Λ and Λ^{*} are one-to-one functions, with each being the inverse of the other. Our second result will assume only that Λ is a function. We state these assumptions below. Let $\mathcal{Y}$ denote the support of Y and $\mathcal{Z}$ denote the support of Z.

Assumption 2 (i) For any given $z \in \mathcal{Z}$, $\Lambda\left(z\right)$ is a singleton; (ii) for any given e, $\Lambda^{*}\left(e\right)$ is a singleton.

Assumption 2(i) is a single mode restriction: For each z, there exists a unique solution e to (3). Note that (i) allows Z to possess a dimension larger than that of ε. Assumption 2(ii) impose further restrictions on the mode of $\varepsilon|Z = z$ requiring that a given mode is only achieved for one single value of z. Assumption 2(i) and (ii) together imply that $\Lambda\left(z\right)$ is one-to-one and so restrict the dimensions of Z and ε to be equal. We note, however, that Z can be a vector of indices that aggregate the effect of a larger vector, through indices that can be estimated in a first step. (See Section 7.2) The following assumption restricts the stochastic relationship between ε and X:

.

Assumption 3 ε is such that $\varepsilon|\left(X, Z\right) \stackrel{d}{=} \varepsilon|Z$ where $\varepsilon|Z$ has a continuous distribution characterized by a density $f_{\varepsilon|Z}\left(e|z\right)$, which is twice continuously differentiable with respect to (z, e).

Assumption 3 restricts ε and Z to be continuous, while X can follow any distribution, including a fully discrete one. We consider the case where Z is discrete in Appendix C. The assumption specifies that X is exogenous, given Z. However, this is not a critical condition for our results to go through: In Section 4.1, we show how our identification argument is easily modified to allow for endogenous regressors.

Assumptions 1 and 3 imply that $Y|\,(X,Z)=(x,z)$ is continuously distributed with density

$$f_{Y|X,Z}(y|x,z) = f_{\varepsilon|Z}\left(r\left(y,x\right)|z\right)\left|\frac{\partial r\left(y,x\right)}{\partial y}\right|, \tag{6}$$

where $|\partial r\left(y,x\right)/\partial y|$ denotes the absolute value of the Jacobian determinant of $r\left(y,x\right)$ with respect to y. This equation establishes a relationship between the conditional density of the observed vector Y at $Y=y$ and the conditional density of the unobserved vector ε when the value of ε is the one satisfying $y=m\left(x,\varepsilon\right)$. Importantly, because of the external nature of Z, the Jacobian determinant does not depend on Z. Taking logs on both sides of this equation and differentiating each with respect to z, gives

$$\frac{\partial \log f_{Y|X,Z}(y|x,z)}{\partial z} = \frac{\partial \log f_{\varepsilon|Z}\left(r\left(y,x\right)|z\right)}{\partial z}. \tag{7}$$

This expression implies that, given y and x, the value z^* of Z such that $\partial \log f_{Y|X,Z}(y|x,z^*)/\partial z = 0$ is also the value of Z such that $\partial \log f_{\varepsilon|Z}\left(r\left(y,x\right)|z^*\right)/\partial z = 0$. Since $r\left(y,x\right)$ is the value e of ε satisfying $y=m\left(x,e\right)$, the value z^* of Z at which

$$\frac{\partial \log f_{Y|X,Z}(y|x,z^*)}{\partial z} = 0 \tag{8}$$

is, by Assumption 2, the value of Z at which such e is most likely. The value z^* is unique due to Assumption 2(ii). When the value of X changes from x to x', while the value e of ε stays fixed, the response y' such that $y'=m\left(x',e\right)$ must be such that

$$e = r\left(y,x\right) = r\left(y',x'\right).$$

But then, y' must also satisfy

$$\frac{\partial \log f_{\varepsilon|Z}\left(r\left(y',x'\right)|z^*\right)}{\partial z} = \frac{\partial \log f_{\varepsilon|Z}\left(e|z^*\right)}{\partial z} = \frac{\partial \log f_{\varepsilon|Z}\left(r\left(y,x\right)|z^*\right)}{\partial z} = 0.$$

By (7), y' is then the solution to

$$\frac{\partial \log f_{Y|X,Z}(y'|x',z^*)}{\partial z} = 0 \tag{9}$$

where x' is given and z^* is the solution to (8). The value y' is also unique due to Assumption 2(i).

These arguments provide a way to constructively identify y' given y, when it is known that for some unknown function m and for some unobservable value, e, of the vector ε, $y = m(x, e)$ and $y' = m(x', e)$: First, find z^* such that eq. (8) is satisfied. Next, given z^*, find y' such that eq. (9) holds. We state this result in the next theorem:

Theorem 1 *Suppose that Assumptions 1-3 are satisfied. For any $y \in \mathcal{Y}$ and $x, x' \in \mathcal{X}$, let e and y' be such that $y = m(x, e)$ and $y' = m(x', e)$. Then, y' is point identified by eqs. (8)-(9) and therefore $\widetilde{\Delta}_y(x, x') = y' - y$ is also point identified.*

In Theorem 1, Assumption 2 makes it possible to establish the existence of a unique z^* that would make the value e of ε the most likely for an observed pair (y, x). Such uniqueness would in general not be possible if Z had different dimension than that of ε. However, as long as for a value z of Z, the value of ε satisfying (3) is unique, that is, when Assumption 2(i) holds, we can still trace out the values of Y across different values of X making sure that the value e of ε stays fixed and satisfies $y' = m(x', e)$. We establish such result next, and provide its proof in the Appendix. In Subsection 7.2, we consider further the case where the dimension of Z is larger than that of ε, by obtaining an estimator for Λ that satisfies Assumption 2.

Theorem 2 *Suppose that Assumptions 1, 2(i) and 3 are satisfied. For any $x, x' \in \mathcal{X}$ and $z \in \mathcal{Z}$, the value of $m(x, \Lambda(z))$ is point identified and therefore $\Delta_{\Lambda(z)}(x, x') = m(x', \Lambda(z)) - m(x, \Lambda(z))$ is also point identified.*

Theorem 2 provides a weaker identification result since $\Lambda(z)$ is not identified. Thus, without further restrictions, it does not allow us to identify the particular individual in the population whose counterfactual response we are measuring. However, in some situations, we might be interested in the value of ε, or on the response of the vector function $m(x, \cdot)$ to changes in the value of ε for any given x. These can be identified if Λ is specified, as assumed next.

Assumption 4 $\Lambda : \mathcal{Z} \to \mathcal{E}$ is known and onto $\mathcal{E}$.

One can interpret Assumption 4 as a normalization analogous to setting $E[\varepsilon|X] = 0$ in an additive regression model, $Y = m(X) + \varepsilon$. In such models, the effect of a change in the value a regressor is identified irrespective of the specific value one imposes on $E[\varepsilon|X]$ as long as it is constant. But the value of the constant term, and therefore the value of the function for any value of the unobservable, cannot be identified without specifying the value of the conditional expectation of the unobservable. Assumption 4 plays this role in our nonseparable model. Under this assumption, the following result follows immediately from our previous results.

Theorem 3 *Suppose that Assumptions 1-4 are satisfied. Then $m(x, e)$ is point identified and therefore $\Delta_e(x, x')$ is also point identified for any given $x, x' \in \mathcal{X}$ and any $e \in \mathcal{E}$.*

4 Extensions and Discussion

In this section, we extend the results in several directions and provide additional interpretations of our assumptions. We first consider the case where the vector of unobservables ε in not distributed independently of X, conditional on Z. We describe how to deal with this situation using a triangular structure. We next consider the case where interest lies on the identification of the derivatives of the structural equations generating the response function. We show how our result can be used to guarantee that the value of ε stays constant when identifying those derivatives across different values of X. In the following subsection, we provide results that relax the assumptions made in Section 3 requiring that $\Lambda(z)$ and $\Lambda^*(z)$ be singletons. We establish partial identification results for the cases where $\Lambda(z)$ and $\Lambda^*(z)$ are set-valued. In the last subsection, we discuss further the interpretation and properties of our external variables.

4.1 Endogenous regressors

Assumption 3 restricts the regressors X to be (conditionally) exogenous. We here remove this assumption by extending our model so that it takes the form of a triangular system. We then use the so–called control function approach to show identification as in Imbens and Newey (2009) and Blundell and Matzkin (2014).

We assume that $X = (X_1, W_1)$ where X_1 and W_1 are vectors of endogenous and exogenous variables, respectively. We then assume that the endogenous variables satisfy

$$X_1 = \pi(W, \eta)$$

where $W = (W_1, W_2) \in \mathcal{W}$ with W_2 being another observed vector and $\eta \in \mathcal{T}$ is an unobserved error component which is independent of W conditional on Z. However, we allow for η and ε to be dependent in which case X_1 is endogenous. The idea is now to use (a transformation of) η as control variable for the endogenous component. To this end, for any given $t \in \mathcal{T}$, define

$$\Lambda(z,t) := \left\{ e \in \mathcal{E} : \frac{\partial f_{\varepsilon|Z,\eta}(e|z,t)}{\partial z} = 0 \right\}, \tag{10}$$

and

$$\Lambda^*(e,t) := \left\{ z \in \mathcal{Z} : \frac{\partial f_{\varepsilon|Z,\eta}(e|z,t)}{\partial z} = 0 \right\}. \tag{11}$$

We then replace Assumptions 2–3 with:

Assumption 2* (i) For any given $(z,t) \in \mathcal{Z} \times \mathcal{T}$, $\Lambda(z,t)$ is a singleton; (ii) for any given e, $\Lambda^*(e,t)$ is a singleton.

Assumption 3* (i) The function $\pi(w,t)$ is invertible in t for all $w \in \mathcal{W}$; (ii) $\pi(w,t)$ is identified up to some unknown one-to-one transformation T, that is, $(w,t) \mapsto \pi(w, T(t))$ is identified; (iii) (ε, η) and W are mutually independent conditional on Z and (iv) $\varepsilon|(Z,\eta)$ has a continuous distribution with the density $f_{\varepsilon|Z,\eta}(e|z,t)$ being twice continuously differentiable with respect to (z,e).

The discussion of Assumption 2 carries over to Assumption 2* with obvious modifications. Assumption 3*(i)-(ii) allows us to identify $T^{-1}(\eta)$ which in turn can be used as a control variable: Under (iii), X and ε are mutually independent conditional on $\left(Z, T^{-1}(\eta)\right)$. Our identification argument now proceeds as before, except that we now throughout condition on $\left(Z, T^{-1}(\eta)\right)$, and we obtain:

Theorem 4 *Suppose that Assumptions 1, 2* and 3* are satisfied. Then the claims of Theorems 1–2 remain true. If in addition Λ as given in (10) is known then the claims of Theorem 3 also hold true.*

4.2 Structural derivatives in models with simultaneity

In this subsection, we provide a new result for identification of changes in the derivatives of structural functions when the value of conditional variables change while those of the unobservable variables stay fixed. As noted in Section 2, identification of functions and derivatives in nonparametric simultaneous equations models has been studied using a transformation of variables equation in several previous works. Such studies imposed additivity and independence restrictions. In addition, some of those methods, as those in Matzkin (2015), are not appropriate when X is discrete for analyzing the effect of changes in X when the value of ε stays fixed.[3] Our method can be used when X is either discrete of continuously distributed and it does not require the additivity and independence assumptions assumed in all those previous works.

For any $j \in \{1, ..., J\}$, we will denote by Y_{-j} the vector Y without the j-th coordinate. We consider the often specified system of simultaneous equations

$$Y_j = s^j \left(Y_{-j}, X, \varepsilon_j\right) \qquad\qquad j = 1, .., J$$

We denote by

$$Y = m\left(X, \varepsilon\right)$$

the system of reduced form functions generated from the above systems of simultaneous equations. Identification of elements in a system of simultaneous equation as above often requires exclusion restrictions on the regressors X, in addition to the exclusion restrictions on ε. Our method achieves identification through Z instead of through X. We next provide conditions for the identification of

$$\frac{\partial s^j \left(y'_{-j}, x', e_j\right)}{\partial y_{-j}} - \frac{\partial s^j \left(y_{-j}, x, e_j\right)}{\partial y_{-j}}$$

for $j \in \{1, ..., J\}$ at the value $\varepsilon = e$ at which $y = m\left(x, e\right)$ and $y' = m\left(x', e\right)$. When $y' = y$, this difference is a discrete version of "cross-partial" derivatives of the structural function. In a demand and supply example, where $J = 2$, s^1 and s^2 could denote, respectively, the demand and supply functions, with Y_1 denoting quantity, Y_2 denoting price, X denoting market characteristics, and ε_1 and ε_2 denoting respectively unobserved taste and unobserved productivity. If $y = m\left(x, e\right)$, then

[3] Matzkin (2015) provides several estimation methods for derivatives of structual functions in simultaneous equations models. These can be calculated at any value of a discrete conditioning variable. Fixing the value of the structural functions at one point of the continuosly distributed variables, and integrating the derivatives, the functions and therefore the value of ε can be identified. However, when a coordinate is discrete, this procedure requires fixing the value of the structural function at one point of the continuous variables *for each value* of the discrete variable. Hence, being able to guarantee that the value of ε stays fixed across different values of the discrete coordinate depends on the arbitrary choices for the function's values.

y is the vector of quantity and price of equilibrium when markets characteristics equal x and the vector of unobserved tastes and productivity equals e. When $y' = y$, our result below establishes identification of the changes in the slopes of the demand and supply functions when the market characteristics change while unobserved taste and productivity stay fixed.

We will make assumptions 1-3 and the following

Assumption 5 For $z \in Z$ such that $e \in \Lambda(z)$, and for $j, k \in \{1, ..., J\}$ such that $k \neq j$

$$\frac{\partial f_{\varepsilon|Z}(e|z)}{\partial z_j \partial \varepsilon_j} \neq 0 \text{ and } \frac{\partial f_{\varepsilon|Z}(e|z)}{\partial z_j \partial \varepsilon_k} = 0$$

This assumption implicitly assigns a coordinate of z to the $j-th$ equation. Such assignment is not necessary to identify the counterfactual y'. It is only necessary if interest lies on the identification of the structural rather than the reduced form functions. Our result is stated in the following theorem.

Theorem 5 *Suppose that Assumptions 1, 2, 3, and 5 are satisfied. If $y \in \mathcal{Y}$, $x, x' \in \mathcal{X}$, and e are such that $y = m(x, e)$, then, for $y' \in \mathcal{Y}$ such that*

$$y' = m(x', e)$$

the structural derivatives

$$\frac{\partial s^j \left(y'_{-j}, x', e_j\right)}{\partial y_{-j}} \text{ and } \frac{\partial s^j \left(y_{-j}, x, e_j\right)}{\partial y_{-j}}$$

are identified.

4.3 Partial Identification Results

Next, we show how Assumption 2 can be relaxed. Assumption 2(i) requires the existence of a unique mode for each z, while Assumption 2(ii) requires that for a given value of e at which the density $f_{\varepsilon|Z}(e|z)$ has a mode, there is a corresponding unique value of z. If either of the two parts is violated, we can only establish partial identification of the counterfactual $\widetilde{\Delta}_y(x, x')$, as stated in the following result.

Theorem 6 *Suppose that Assumptions 1, 2(ii), and 3 are satisfied. For any $y \in \mathcal{Y}$ and $x, x' \in \mathcal{X}$, let z^* be the unique solution to (8). Then,*

$$\widetilde{\Delta}_y(x, x') \in \left\{ y' - y \in \mathbb{R}^{d_Y} : \frac{\partial \log f_{Y|X,Z}(y'|x', z^*)}{\partial z} = 0 \right\}.$$

Suppose that Assumptions 1 and 3 are satisfied. For any $y \in \mathcal{Y}$ and $x, x' \in \mathcal{X}$,

$$\widetilde{\Delta}_y\left(x, x'\right) \in \left\{ y' - y \in \mathbb{R}^{d_Y} : \text{ For some } z^* \in Z, \quad \frac{\partial \log f_{Y|X,Z}\left(y|x, z^*\right)}{\partial z} = 0 \quad \& \quad \frac{\partial \log f_{Y|X,Z}\left(y'|x', z^*\right)}{\partial z} = 0 \right\}.$$

When the focus is on counterfactuals that do not specify the value of ε other than the requirement that it stays constant, the partial identification version of Theorem 2 is:

Theorem 7 *Suppose that Assumptions 1 and 3 are satisfied. For any $x, x' \in \mathcal{X}$ and $z \in \mathcal{Z}$,*

$$\Delta_{\Lambda(z)}\left(x, x'\right) \in \left\{ y' - y \in \mathbb{R}^{d_Y} : \frac{\partial \log f_{Y|X,Z}\left(y|x, z\right)}{\partial z} = 0 \quad \& \quad \frac{\partial \log f_{Y|X,Z}\left(y'|x', z\right)}{\partial z} = 0 \right\}.$$

4.4 The ε-Z relationship

The strength of our identification results depend on the features of $\Lambda(z)$ defined in (4) If this is one-to-one then point identification is achieved – otherwise, only partial identification is achieved. We here discuss the properties of Λ in a given application and how one should choose the external regressors in Z to ensure point identification. Before doing so, it should be noted that the assumptions made on the mappings Λ and Λ^* can be tested by exploiting eq. (7), since the properties of Λ and Λ^* are embedded in $\partial f_{Y|X,Z}\left(y|x, z\right) / \left(\partial z\right)$. For example, invertibility of Λ is guaranteed by invertibility of $\partial f_{Y|X,Z}\left(y|x, z\right) / \left(\partial z\right)$ w.r.t. z. Thus, point identification is testable. We utilize this fact in Section 7 where we develop data-driven selection methods for Z.

How should we then choose Z to obtain point identification? Recall that the relationship between ε and Z determines the properties of the solution mappings, (4) and (5). At one extreme, if there exists a stochastic one-to-one mapping between the two, $\Lambda(z)$ will generally be an invertible real-valued function, see below. In order for this to hold, Z must necessarily be of at least the same dimension as ε, $d_Z \geq d_\varepsilon$ and Z and ε must covary. At the other extreme, suppose that Z and ε are fully independent so no stochastic relationship exists between the two. Then $f_{\varepsilon|Z}\left(e|z\right) = f_\varepsilon\left(e\right)$ and $\Lambda(z) = \mathcal{E}$ which is the maximum volume that it can achieve. Examples 1–2 below illustrate the features of $\Lambda(z)$. Finally, note that $\Lambda(z) = \varnothing$ is ruled out since $f_{\varepsilon|Z}\left(e|z\right)$ is a density.

Thus, the researcher should choose Z to ensure maximal covariation between Z and ε in terms of $f_{\varepsilon|Z}\left(e|z\right)$. In a given application, the choice of variables included in Z should therefore reflect the type of unobserved heterogeneity that enters the model of interest. In the consumer demand example, one could think of each of the components of ε as capturing a particular type of tastes/preferences of the consumer. We then need to identify corresponding socio-economic characteristics in data that we expect are capturing variation in these unobserved tastes.

It is important to stress that Λ does not describe the stochastic relationship between the underlying random variables Z and ε. For example, for a given individual characterized by (ε, Z), it will not hold that $\partial f_{\varepsilon|Z}(\varepsilon|Z)/(\partial z) = 0$, and so we cannot directly use $\partial f_{\varepsilon|Z}(\varepsilon|Z)/(\partial z)$ to identify an individual's particular value of ε. But Λ does provide information about the distributional relationship, and this suffices for our identification results.

Our machinery accommodates for Z to be either a noisy proxy or noisy measurement of ε. The following two examples illustrate these two scenarios:

Example 1 Suppose that Z acts as a noisy *proxy* for ε, so that ε satisfies

$$t(\varepsilon) = s(Z) + \eta,$$

for some unknown functions s and t and some unobservable vector η which is independent of Z. Suppose furthermore that t is invertible in which case

$$\log f_{\varepsilon|Z}(e|z) = \log f_\eta(t(e) - s(z)) + \log|\partial t(e)/\partial e|.$$

and so

$$\frac{\partial \log f_{\varepsilon|Z}(e|z)}{\partial z} = -\frac{\partial s(z)}{\partial z}\frac{\partial \log f_\eta(t(e) - s(z))}{\partial \eta}.$$

Thus,

$$\Lambda(z) = \left\{e \in \mathcal{E} : \frac{\partial s(z)}{\partial z}\frac{\partial \log f_\eta(t(e) - s(z))}{\partial \eta} = 0\right\}$$

which will generally be a set. The properties of s determines how closely ε and Z covary. For example, if $\partial s(z)/(\partial z)$ has full rank and the distribution of η has a unique mode at zero, then $\Lambda(z) = t^{-1}(s(z))$ is a singleton.

Example 2 Suppose instead that Z acts as a noisy and possibly biased *measurement* of ε so that

$$Z = s(\varepsilon, \eta)$$

where s is unknown and η is unobserved. Assume that the marginal distribution of Z is uniform.[4] Then, since $f_{\varepsilon|Z}(e|z) = f_{\varepsilon,Z}(e,z)/f_z(z) = f_{Z|\varepsilon}(z|e)\ f_\varepsilon(e)/f_Z(z)$,

$$\frac{\partial f_{\varepsilon|Z}(e|z)}{\partial z} = \frac{\partial f_{Z|\varepsilon}(z|e)}{\partial z}.$$

Thus, the properties of Λ are determined by the ones of $s(\varepsilon,\eta)$ and η. If, for example, $s(\varepsilon,\eta) = \bar{s}(\varepsilon)+\eta$ for some $\bar{s}(\cdot)$ and the distribution of η has a unique mode at zero; then

$$\frac{\partial f_{\varepsilon|Z}(e|z)}{\partial z} = \frac{\partial f_{Z|\varepsilon}(z|e)}{\partial z} = \frac{\partial f_\eta(z-\bar{s}(e))}{\partial \eta},$$

and so

$$\Lambda(z) = \{e \in \mathcal{E} : z = \bar{s}(e)\},$$

which is a singleton if $\bar{s}$ is invertible. This includes as a special case the standard measurement error model, where $\bar{s}(\varepsilon) = \varepsilon$.

In general, adding more (relevant) external covariates helps in the identification since the "size" of the set of solutions, $\Lambda(z)$, will generally shrink as we add more score equations that have to be satisfied. However, once point identification has been achieved, so that $\Lambda(z)$ is a singleton and $\Lambda(\mathcal{Z}) = \mathcal{E}$, adding more external covariates provides no gain in terms of establishing identification. This is illustrated in the following example:

Example 3 Suppose that ε, which is assumed to be a scalar for notational simplicity, satisfies

$$\varepsilon = \sum_{i=1}^{\bar{d}} s_i(Z_i) + \eta,$$

where $s_i : \mathbb{R} \mapsto \mathbb{R}$ are one-to-one, $i = 1, ...d$, and, as before, $Z_1, ..., Z_{\bar{d}}$ and η are mutually independent and with full support. We can then normalize these such that $s_i(Z_i) \sim N(0,1)$, $i = 1, ...d$, and $\eta \sim N(0,1)$. It is now easily checked that using the first $d_Z \leq \bar{d}$ external covariates yields the following solution mapping,

$$\Lambda_{d_Z}(z_1, ...z_{d_Z}) = \sum_{i=1}^{d_Z} s_i(z_i).$$

[4]This is without loss of generality since we can always transform Z. And the identifying power of Λ is invariant to invertible transformations of ε and Z: For any two invertible transformations G_ε and G_Z, the conditional distribution of $\bar{\varepsilon} = G_\varepsilon(\varepsilon)\,|\bar{Z} = G_Z(Z)$ satisfies

$$f_{\bar{\varepsilon}|\bar{Z}}(z|\bar{z}) = \frac{f_{\bar{\varepsilon},\bar{Z}}(z,\bar{z})}{f_{\bar{Z}}(\bar{z})} = \frac{f_{\varepsilon,Z}\left(G_\varepsilon^{-1}(z), G_Z^{-1}(\bar{z})\right)}{f_Z\left(G_Z^{-1}(\bar{z})\right)}\left|\frac{\partial G_\varepsilon^{-1}(z)}{\partial z}\right| = f_{\varepsilon|Z}\left(G_\varepsilon^{-1}(z), G_Z^{-1}(\bar{z})\right)\left|\frac{\partial G_\varepsilon^{-1}(z)}{\partial z}\right|$$

and so the solutions mappings for the score equations of $\bar{\varepsilon}|\bar{Z}$, $\bar{\Lambda}$ satisfies $\bar{\Lambda}(\bar{z}) = G_\varepsilon\left(\Lambda\left(G_Z^{-1}(z)\right)\right)$.

Thus, Λ_{d_Z} is a singleton for all choices of $d_Z \geq 1$ and so nothing is gained, in terms of identification of m, from using more external covariates in this case.

Maintain the above model but suppose now that $s_1(\mathcal{Z}_1) = [0, +\infty)$ and $s_2(\mathcal{Z}_2) = (-\infty, 0)$. In this case, using Z_1 alone as external covariate will only allow us to identify individuals with positive values of ε, while using both Z_1 and Z_2 allow us to "hit" all individuals in the population.

5 Application to consumer demand

In our empirical application, we apply our methodology to the estimation of household demand when unobserved tastes stay fixed and prices or income change either in a discrete or a continuous form. Such demand functions can be used to recover individual preferences using revealed preference methods when changes in prices and/or income are discrete, or using either integrability or revealed preference methods when prices and/or income are continuously distributed. Our method allows to identify such individual demand functions even when each individual is observed only once. We here demonstrate how our general framework accommodates a very flexible consumer demand model and how our identification argument works in this context.

Consider a consumer characterized by income level $I \in \mathbb{R}_+$ (representing the total budget available to the consumer potentially adjusted for the set of consumer goods of interest) together with unobserved individual characteristics which we collect in $\varepsilon \in \mathcal{E}$. The consumer chooses quantities of $d_Y + 1$ divisible goods. Let $p = (p_1, ..., p_{d_Y})' \in \mathbb{R}_+^{d_Y}$ denote the (relative) prices of the first d_Y goods, where we leave out the last good whose demand is identified through the budget constraint. Given these prices, the consumer demands $Y = (Y_1, ..., Y_{d_Y})' \in \mathcal{Y} \subseteq \mathbb{R}_+^{d_Y}$. We let m denote the demand function that maps prices, income and consumer characteristics into demands

$$Y = m(p, I, \varepsilon). \tag{12}$$

A parametric approach would deal with unobserved heterogeneity by imposing a specific functional form. Consider, for example, an extended CES utility specification for a consumer with unobserved tastes $\varepsilon_1, ..., \varepsilon_G$ given by

$$U(Y, \varepsilon) = (\sum_{g=0}^{d_Y} \alpha_g Y_g^{\theta})^{1/\theta} + \sum_{g=1}^{d_Y} \varepsilon_g Y_g^{\rho_g},$$

where Y_0 is the "numeraire good", $0 < \theta < 1$ is the CES parameter, and $0 < \rho_g < 1$ for $g = 1, ..., G$. As can be verified either directly or by applying Theorem 8 below, the corresponding system of demand functions is invertible in ε and hence we can write

$$\varepsilon = r(p, I, Y). \tag{13}$$

That is, given any value of $Y = (Y_1, ..., Y_{d_Y})$ one can pin down the value of $\varepsilon = (\varepsilon_1, ..., \varepsilon_{d_Y})$ that generated it. It is then easy to predict the demand for a consumer with the same ε when confronted with a different budget set. This analysis, however, entirely rests on the strong parametric specification of preferences being correct. If incorrect, the analysis will be invalid. It seems plausible that the underlying utility function has a more flexible structure than the one above. In this case the consumer's choice for one commodity depends on the unobservable tastes for all commodities, and pinning down the value of ε is much more challenging. Our identification result allows us to do so in fairly straightforward manner.

A primary goal of much consumer demand analysis is to measure the impact of changes in prices p and income levels I on the demand. With the identification result developed in the previous sections, this can be achieved at an individual level. Consider a consumer characterized by $\varepsilon = e$ who has income $I = i_0$ and faces a price shock changing prices $p = p_0$ to p_1. We might think here of a low income consumer (low i_0) with a high initial consumption of some key goods $y_0 = m(p_0, i_0, e)$. The counterfactual demand response for consumer (y_0, i_0, e) is given by

$$\Delta_e = m(p_1, i_0, e) - m(p_0, i_0, e). \tag{14}$$

Our identification argument will allow us to identify Δ_e. The value of ε at which the counterfactuals will be identified can be defined from the initial observed demand $y_0 = m(p_0, i_0, e)$. The counterfactual response, when ε is so characterized, is identified without a normalization. Adding a normalization, to assign numerical values to ε, we will be able to identify changes in m with respect to e, and also identify the distribution of ε. Our identification approach will also allow us to identify marginal effects such as $\partial m(p, I, e) / (\partial I)$ when I is continuously distributed and $\partial m(p, I, e) / (\partial p)$ when p is continuously distributed.

The above demand analysis is feasible within our framework under weak additional regularity conditions on the demand model. Consider first Assumption 3: As mentioned in Section 1, several recent results exist on invertibility of demand functions, such as Brown and Matzkin (1998), Beckert and Blundell (2008), Berry, Gandhi and Haile (2013) and Chiappori et al (2016a). The following result is a variation on the results in Brown and Matzkin (1998) and Beckert and Blundell (2008).

It provides a practical method for incorporating multidimensional unobserved heterogeneity around commonly used deterministic utility functions, in a way that generates invertible demand function. We suppress any dependence on observables W since they remain fixed, and let $y = (y_1, ..., y_{d_Y})$ denote the demand of the first d_Y goods.

Theorem 8 *Suppose that the utility function satisfies* $U(y, y_{d_Y+1}, \varepsilon) = U_1(y, y_{d_Y+1}) + U_2(y, \varepsilon)$ *where: (i)* $U_1(y, y_{d_Y+1})$ *is a twice continuously differentiable, strictly increasing and strictly quasiconcave function; (ii)* $U_2(y, \varepsilon)$ *is a twice continuously differentiable function, which for each* ε *is strictly increasing and strictly concave in y; (iii) for any y, all the principal minors of the matrix* $D_{y\varepsilon}U_2(y, \varepsilon) = \left[\partial^2 U_2(y, \varepsilon) / (\partial y_i \partial \varepsilon_j)\right]_{i,j=1}^{d_Y}$ *are strictly positive. Then, the (demand) function* $y = m(p, I, \varepsilon)$ *that maximizes* $U(y, y_{d_Y+1}, \varepsilon)$ *subject to the budget constraint* $p'y + y_{d_Y+1} \leq I$ *is invertible in* ε.

Assumption 1 requires that, in addition to (p, I), we have observed a set of consumer-specific covariates Z which covary with ε and, conditional on ε, do not enter directly into the demand function (12). Suppose that we observe a number of individual characteristics for each consumer. These will comprise two groups: The first group of characteristics, W, is included in $X = (p, I, W)$ and so we control for the effects of these on demands explicitly. The second group of characteristics is included in Z to be used as external covariates. The second group of observed characteristics is in this sense absorbed into the unobserved component ε, and so we do not control for the effect of the second set on demand explicitly. In our empirical application, Z is computed as an index of household members' attitudes and behaviours relating to consumer preferences, for example the type of news papers and TV stations they see, their attitudes to fitness, their education, etc; it seems plausible that these variables affect consumer preferences.

One will in general expect income I (total budget to the subset of goods) to be endogenous, and that certain observed characteristics comove with the unobserved components. Here, we may introduce instruments W and use a control function approach to control for endogeneity as discussed earlier.

6 Nonparametric Estimation and Inference

We develop in this section nonparametric estimators of $\tilde{\Delta}_y(x, x')$ and $\Delta_e(x, x')$ based on the identification results in the previous section. We will in the following assume that X has a continuous distribution – the case of discrete regressors is easily accommodated for but the notation gets more cumbersome. Moreover, in our theoretical analysis, we restrict ourselves to the case where Assumption 2 is satisfied and so Λ is a function.

First, we define the relevant objects to be estimated. Recall that the population version of the counterfactual effect was identified through the solutions z^* to eq. (8) and the solution y' to eq. (9). Note that these are actually functions and so we will use the following notation in this section: For any given values of (y,x), let $\bar{r}(x,y)$ be the solution to

$$\frac{\partial \log f_{Y|X,Z}(y|x,\bar{r}(x,y))}{\partial z} = 0. \tag{15}$$

Thus, $\bar{r}(x,y)$ corresponds to z^* where we now emphasize that it depends on (x,y). Similarly, for any given value of (x,z), we let $\bar{m}(x,z)$ be the solution to

$$\frac{\partial \log f_{Y|X,Z}(\bar{m}(x,z)|x,z)}{\partial z} = 0. \tag{16}$$

This corresponds to y'. In terms of these two functions, the counterfactual effect is given by

$$\tilde{\Delta}_y(x,x') = \bar{m}(x,\bar{r}(x,y)) - y. \tag{17}$$

Similarly, for a given z, we can represent $\Delta_{\Delta(z)}(x,x')$ as

$$\Delta_{\Delta(z)}(x,x') = \bar{m}(x',z) - \bar{m}(x,z). \tag{18}$$

Let (Y_i, X_i, Z_i), $i = 1, ..., n$, be i.i.d. observations from the model. We then propose the following GMM-type estimators of $\bar{m}(x,z)$ for any given values of $(x,z) \in \mathcal{X} \times \mathcal{Z}$:

$$\hat{m}(x,z) = \arg\min_{y \in \mathcal{Y}_0} \|\hat{g}(y|x,z)\|^2, \tag{19}$$

where

$$\hat{g}(y|x,z) = \frac{\partial \hat{f}_{Y|X,Z}(y|x,z)}{\partial z'} \tag{20}$$

contain the "moment" conditions, $\hat{f}_{Y|X,Z}$ is a nonparametric estimator of $f_{Y|X,Z}$ and $\mathcal{Y}_0 \subseteq \mathcal{Y}$ is some compact subset that the true function value $\bar{m}(x,z)$ is assumed to lie in. Ideally we would like to set $\mathcal{Y}_0 = \mathcal{Y}$, but, as with other extremum estimators whose objective function is potentially non-convex, we have to restrict the set of candidate values to be compact. Similarly, an estimator of $\bar{r}(y,x)$ can be obtained by either computing

$$\hat{r}(y,x) = \arg\min_{z \in \mathcal{Z}_0} \|\hat{g}(y|x,z)\|^2, \tag{21}$$

or

$$\tilde{r}(y,x) = \arg\max_{z \in \mathcal{Z}_0} \hat{f}_{Y|X,Z}(y|x,z), \tag{22}$$

where $\mathcal{Z}_0 \subseteq \mathcal{Z}$ is some compact subset that the true function value $\bar{r}(y, x)$ is assumed to lie in.

Any nonparametric conditional density estimator could in principle be employed in the implementation of the above estimators. We here focus on the case where $\hat{f}_{Y|X,Z}$ has been chosen as a kernel density estimator, of the form

$$\hat{f}_{Y|X,Z}(y|x,z) = \frac{\sum_{i=1}^{n} K_{Y,h_Y}(Y_i - y) K_{X,h_X}(X_i - x) K_{Z,h_Z}(Z_i - z)}{\sum_{i=1}^{n} K_{X,h_X}(X_i - x) K_{Z,h_Z}(Z_i - z)}, \tag{23}$$

where $K_{a,h_a} = K_a(\cdot/h_a)/h_a$, $K_a : \mathbb{R}^{d_a} \mapsto \mathbb{R}$ is a kernel function, and $h_a > 0$ a bandwidth, $a \in \{Y, X, Z\}$. If X has discrete components, $K_{X,h_X}(X_i - x)$ in the above expression should be replaced by $K_{X,h_X}(X_{1,i} - x_1)\mathbb{I}\{X_{2,i} = x_2\}$ where X_1 and X_2 contain the continuous and discrete components of X, respectively, and $\mathbb{I}\{\cdot\}$ denotes the indicator function. With this modification of the estimator, all the following asymptotic statements remain correct for the mixed discrete-continuous case as well by letting d_X denote the dimension of X_1.

For the asymptotic analysis of $\hat{m}$, we impose regularity conditions on the kernel functions used to compute $\hat{f}_{Y|X,Z}$, and the underlying structure of the model at the values (x, z) at which we wish to estimate $\bar{m}$. To state the conditions, we introduce

$$H(x,z) = \left.\frac{\partial^2 f_{Y|X,Z}(y|x,z)}{\partial z \partial y'}\right|_{y=\bar{m}(x,z)} \in \mathbb{R}^{d_Y \times d_Y}, \tag{24}$$

which measures the information content of the "moment" conditions.

Assumption 5 The kernel functions are twice continuously differentiable and satisfy the following conditions: $\int_{\mathbb{R}^{d_a}} K_a(x)\,dx = 1$, $\int_{\mathbb{R}^{d_a}} xK_a(x)\,dx = 0$ and $\int_{\mathbb{R}^{d_a}} \|x\|^2 K_a(x)\,dx < \infty$ for $a \in \{Y, X, Z\}$.

Assumption 6 (i) The function $m(x, e)$ is twice continuously differentiable w.r.t. e; (ii) (X, Z) has a continuous distribution whose density, $f_{X,Z}(x, z)$ is twice continuously differentiable with $f_{X,Z}(x, z) > 0$.

Assumption 7 (i) $\bar{m}(x, z)$ is situated in the interior of $\mathcal{Y}_0$; (ii) $H_m(x, z)$ defined in eq. (24) has full rank.

Assumptions 5–6 allow us to apply standard results from the analysis of nonparametric kernel estimators. In particular, Assumptions 1, 3 and 6 guarantee that the joint density of (Y, X, Z), $f_{Y,X,Z}(y, x, z) = f_{Y|X,Z}(y|x, z) f_{X,Z}(x, z)$, exists and is twice continuously differentiable. This combined with the use of second-order kernels, as imposed in Assumption 5, imply that the leading bias terms of $\partial\hat{f}_{Y|X,Z}(y|x,z)/(\partial z')$ are of order $O_P(h_Y^2) + O_P(h_X^2) + O_P(h_Z^2)$, while the variance terms are of order $O_P\left(1/[nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Y+2}]\right)$. The overall bias could be reduced by using higher-order kernels combined with assuming the existence of higher-order derivatives of m and $f_{\varepsilon,Z}$; however, to avoid overly complicated assumptions, we refrain from this here.

Assumption 6 is used for the analysis of the GMM-type estimator $\hat{m}(x,z)$ and contains standard conditions found in the analysis of GMM estimators: Assumption 7(i) together with the identification result in Theorem 8 ensure that the GMM estimator defined in eq. (19) is consistent; Assumption 7(ii) rules out that the "true" parameter lies on the boundary of the parameter space; and Assumption 7(iii) is the usual rank condition for GMM estimators that guarantee local identification.

The analysis of the estimators follow along the same lines as the one for standard GMM estimators with the exception that the sample moment conditions here takes the form of the first-order derivatives of a kernel density estimator. In particular, the convergence rate of $\hat{m}$ will be determined by the ones of the density derivative estimator.

Theorem 9 *Suppose that Assumptions 1–3 and 5–7 hold. Then, for any bandwidth sequences satisfying*

$$nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}h_a^4 \to 0 \text{ for } a = Y, X, Z, \ \ \log(n)/\left(nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}\right) \to 0, \text{ and } nh_Y^{d_Y+2}h_X^{d_X}h_Z^{d_Z+2} \to \infty, \tag{25}$$

the estimator $\hat{m}(x,z)$, as defined by eq. (19), is consistent and satisfies

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}}\{\hat{m}(x,z) - \bar{m}(x,z)\} \to^d N(0, V_m(x,z)),$$

where

$$V_m(x,z) = H_m^{-1}(x,z)\,\Omega_m(x,z)\,H_m^{-1}(x,z),$$

and

$$\Omega_m(x,z) = \left.\frac{f_{Y|X,Z}(y^*|x,z)}{f_{X,Z}(x,z)}\right|_{y^*=\bar{m}(x,z)} \int_{\mathbb{R}^{d_Y}} K_Y^2(y)\,dy \int_{\mathbb{R}^{d_X}} K_X^2(x)\,dx \int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z(z)}{\partial z}\frac{\partial K_Z(z)}{\partial z'}dz \in \mathbb{R}^{d_Y\times d_Y}. \tag{26}$$

Remark 1. The first and second bandwidth condition in eq. (25) control the bias and variance of $\hat{g}(y|x,z)$, respectively, and ensure that they vanish sufficiently fast. The third condition implies that the nonparametric estimator $\hat{H}_m(x,z) = \partial^2\hat{f}_{Y|X,Z}(y^*|x,z)/(\partial z\partial y')$ is consistent

We observe that the usual curse-of-dimensionality of nonparametric estimators is present: The convergence rate of $\hat{m}$ deteriorates as the dimensions of Y, X and/or Z increase. Moreover, given these, the asymptotic variance, $V_m(x,z)$, of $\hat{m}$ takes the usual sandwich form as well-known for

GMM estimators. The over all variance depends on two properties of the model: First, $\Omega_m(x,z)$ is the standard asymptotic variance of kernel density derivatives and so captures the precision with which we can learn about the true density derivative (”moment conditions”). Second, as discussed earlier, $H_m(x,z)$ measures the identifying strength of Z as it measures the local curvature of the first-order conditions identifying $\bar{m}$.

Next, we analyze the two estimators of $\bar{r}(y,x)$ defined in eqs. (21)-(22). We impose the following additional assumption for this analysis, which corresponds to the conditions imposed in Assumption 6 for the estimation of $\bar{m}$:

Assumption 8 (i) $\bar{r}(y,x)$ is situated in the interior of $\mathcal{Z}_0$ and (ii)

$$H_r(y,x) := \left.\frac{\partial^2 f_{Y|X,Z}(y|x,z)}{\partial z \partial z'}\right|_{z=\bar{r}(y,x)} \in \mathbb{R}^{d_Y \times d_Y} \text{ has full rank.}$$

Theorem 10 *Suppose that Assumptions 1–3, 5–6 and 8 hold. Then, for any bandwidth sequences satisfying*

$$nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2} h_a^4 \to 0 \text{ for } a = Y, X, Z, \quad \log n / h_Y^{d_Y} h_X^{d_X} h_{Z|}^{d_Y+4} \to 0, \text{ and } nh_Y^{d_Y+2} h_X^{d_X} h_Z^{d_Z} \to \infty, \quad (27)$$

the estimator $\hat{r}(y,x)$, as defined by eq. (21), is consistent and satisfies

$$\sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}}\{\hat{r}(y,x) - \bar{r}(y,x)\} \to^d N(0, V_r(y,x)),$$

where

$$V_r(x,z) = H_r^{-1}(x,z)\,\Omega_r(x,z)\,H_r^{-1}(x,z),$$

and

$$\Omega_r(y,x) = \left.\frac{f_{Y|X,Z}(y|x,z^*)}{f_{X,Z}(x,z^*)}\right|_{z^*=\bar{r}(y,x)} \int_{\mathbb{R}^{d_Y}} K_Y^2(z)\,dz \int_{\mathbb{R}^{d_X}} K_X^2(x)\,dx \int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z(z)}{\partial z}\frac{\partial K_Z(z)}{\partial z'} dz \in \mathbb{R}^{d_Y \times d_Y}$$

Furthermore, the estimator $\tilde{r}(y,x)$ defined in eq. (22) is first-order equivalent to $\hat{r}(y,x)$.

One concern one may have, given the slow convergence rate reported in the above theorem, is poor finite-sample performance of the proposed estimator. To investigate how well our estimator performs in finite samples, we carried out a simulation study with the results reported in Appendix E. As can be seen from these results our estimator performs well in sample sizes around $n = 2,000$.

Given the estimators of $\bar{r}$ and $\bar{m}$, natural estimators of the counterfactual effects $\tilde{\Delta}_y(x,x')$ and $\Lambda_{\Lambda(z)}(x,x')$ defined in (17) and (18), respectively, are

$$\hat{\Delta}_y\left(x,x'\right) := \hat{m}(x,\hat{e}) - y,$$

and

$$\hat{\Delta}_{\Lambda(z)}\left(x,x'\right) := \hat{m}\left(x',z\right) - \hat{m}(x,z).$$

Theorem 11 *Suppose that Assumptions 1–3 and 5–8 hold together with eq. (25). Then,*

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Y+2}}\{\hat{\Delta}_{\Lambda(z)}\left(x,x'\right)-\Delta_{\Lambda(z)}\left(x,x'\right)\}\to^d N\left(0,V_m\left(x,z\right)+V_m\left(x',z\right)\right),$$

where $V_m(x,z)$ is defined in Theorem 9. If furthermore eq. (27) hold then

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Y+2}}\{\hat{\Delta}_y\left(x,x'\right)-\tilde{\Delta}_y\left(x,x'\right)\}\to^d N\left(0,V_y\left(x,x'\right)\right),$$

where, with H_r and V_r defined in Theorem 10,

$$V_y\left(x,x'\right)=\frac{\partial m(x',z)}{\partial z'}H_r^{-1}\left(y,x\right)V_r\left(y,x\right)H_r^{-1}\left(y,x\right)\frac{\partial m(x',z)}{\partial z'}+V_m\left(x',\bar{r}\left(y,x\right)\right).$$

6.1 Panel data setting

The estimators and their asymptotic theory developed so far assumed a single cross section of data. In our empirical application, we will work with a panel of n households observed over $T\geq 2$ time periods, $(Y_{i,t},X_{i,t},Z_{i,t})$, for $i=1,....,n$ and $t=1,...,T$. The identification result will still go through if we assume that eq. (1) holds for all individuals at each time point t and that our assumptions are satisfied for all $t=1,...,T$.

The estimation procedure remains large unaltered with the only difference being that we now replace the kernel density estimator in eq. (23) with the following version that pools data across the T time periods,

$$\hat{f}_{Y|X,Z}(y|x,z)=\frac{\sum_{i=1}^n\sum_{t=1}^T K_{Y,h_Y}\left(Y_{i,t}-y\right)K_{X,h_X}\left(X_{i,t}-x\right)K_{Z,h_Z}\left(Z_{i,t}-z\right)}{\sum_{i=1}^n\sum_{t=1}^T K_{X,h_X}\left(X_{i,t}-x\right)K_{Z,h_Z}\left(Z_{i,t}-z\right)}. \tag{28}$$

The asymptotic theory also remains largely unchanged. In particular, we do not need to cluster the standard errors at the individual level.

Theorem 12 *Suppose that $\{(Y_{i,t},X_{i,t},Z_{i,t}):i=1,....,n,t=1,...,T\}$ is generated by*

$$Y_{i,t}=m\left(X_{i,t},\varepsilon_{i,t}\right),$$

where: (i) $\{(\varepsilon_{i,t},X_{i,t},Z_{i,t}):t=1,...,T,\}$ is stationary for each $i=1,...,n$ and independently distributed across $i=1,....,n$, and (ii) for any $s\neq t$, $(\varepsilon_{i,s},X_{i,s},Z_{i,s},\varepsilon_{i,t},X_{i,t},Z_{i,t})$ has a well-defined continuous density. Then, under Assumptions 1–3 and 5–8, the asymptotic results of Theorems 9–11 remain true when $\Omega_m(x,z)$ and $\Omega_r(x,y)$ are replaced by $\Omega_m(x,z)/T$ and $\Omega_r(x,z)/T$, respectively.

7 Choosing External Covariates

To achieve point identification we need to find relevant external covariates Z so that Assumption 3 is satisfied. In most applications, the researcher will either have more Z's available that potentially satisfy Assumption 3, or will be uncertain about whether a potential set of candidate variables are valid. We are then interested in selecting a subset of size d_Y from these that satisfies Assumption 3 for the following two reasons: First, the nonparametric estimator $\hat{m}(x,z)$ suffers from a curse-of-dimensionality with the precision deteriorating as d_Z increases, c.f. Theorem 9. Second, if $d_Z > d_Y$, the estimator $\hat{m}(x,z)$ is not invertible in z and so we cannot recover the distribution of ε (up to the transformation Λ).

We here develop methods for identifying a valid set of external covariates. We take as starting point that we have available $d_Z \geq d_Y$ candidate external variables available which we collect in $Z = (Z_1, ..., Z_{d_Z})'$. Two procedures are then developed: The first procedure tests for whether a given subset of the candidate variables are valid. The second procedure considers a more general scenario where some, potentially nonlinear, transformation of the candidate variables constitutes a valid set of external covariates.

In Appendix D, we extend the asymptotic theory to allow for multiple sets of variables that satisfy Assumption 3.

7.1 Testing for Existence of Sufficient External Covariates

In the following, we take Assumptions 1-2 as maintained hypotheses and then wish to Assumption 3. Consider first the case where we have exactly d_Y external covariates whose validity we wish to test. To that end, first observe that

$$\frac{\partial^2 f_{Y|X,Z}(y|x,z)}{\partial z \partial z'} = \frac{\partial^2 f_{\bar{\varepsilon}|z}\left(\bar{r}\left(y,x\right)|z\right)}{\partial z \partial z'} \left|\frac{\partial \bar{r}(y,x)}{\partial y}\right|.$$

Thus, the rank condition imposed on the matrix $H_r(y,x)$ in Theorem 10 is satisfied if and only if $\partial^2 f_{\bar{\varepsilon}|Z}(\bar{\varepsilon}|z^*)/(\partial z \partial z')$ has full rank. This in turn holds if and only if $\partial^2 f_{\bar{\varepsilon}|Z}(\varepsilon|z)/(\partial z \partial z')$ has full rank and is implied by Assumption 3. Thus, we can test Assumption 3 by testing whether the rank of $H_r(y,x)$ is d_Y or not. The matrix $H_r(y,x)$ can be estimated using standard methods and a rank test for it can be performed using existing tests; see Al-Sadoon (2015) for an overview of such methods and some recent developments.

If we have more than d_Y external covariates ($d_Z > d_Y$), one can test Assumption 3(i) through a nonparametric version of the J-test used in GMM with over-identifying moment conditions: Choose the weighting matrix such that $\hat{W}(x,z) \rightarrow^P \Omega_m^{-1}(x,z)$, where $\Omega_m(x,z)$ is defined in Theorem 16. It now follows from the limit results derived in the proof of Theorem 16 in conjunction with the arguments in Newey and McFadden (1994, Section 9.5) that

$$\hat{J}(x,z) := n h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} \min_{y \in \mathcal{Y}_0} \hat{g}(y|x,z) \hat{\Omega}_m^{-1}(x,z)\, \hat{g}(y|x,z) \rightarrow^d J(x,z) \tag{29}$$

for all (x, z) under Assumption 3(i), where $J(x, z) \sim \chi^2_{d_Z - d_Y}$ and $J(x_1, z_1) \perp J(x_2, z_2)$ for any two pairs (x_1, z_1) and (x_2, z_2). Under the alternative, so that Assumption 3(i) does not hold for some (x, z), we have $\hat{J}(x, z) \to^P +\infty$.

The above two testing procedures can be used to identify external covariates that satisfy Assumption 3: Suppose we have available $d_Z \geq d_Y$ candidate variables of which d_Y satisfies Assumption 3. We can then either directly test for whether a given subset of size d_Y satisfies Assumption 3; or we can use a step-down procedure where one variable at a time is removed according to whether the J statistic of the "reduced" model does not reject the null.

7.2 Constructing an index Z

Instead of selecting a subset from the d_Z candidate variables, one can try to combine all the variables to construct an index Z^* from them. This can be done in a number of ways. At the most general level, the hypothesis of interest can be stated as

$$Z^* = B(Z) \text{ satisfies Assumption 3 for some function } B : \mathbb{R}^{d_Z} \mapsto \mathbb{R}^{d_Y}.$$

A natural way to estimate B is by searching across a set of functions and choose the one that provides the best fit in terms of explaining the variation in Y conditional on X. Suppose that $B \in \mathcal{B}$ for some function space $\mathcal{B}$; we can then combine our kernel density estimator with sieve methods (see Chen, 2007) to estimate B by

$$\hat{B} = \arg \max_{\substack{B \in \mathcal{B}_n \\ \|B\|=1}} \sum_{i=1}^{n} \log \hat{f}_{Y|X,Z^*}(Y_i | X_i, Z_i; B),$$

where $\mathcal{B}_n$, $n \geq 1$, is a sequence of approximating parameter spaces (sieves) that becomes dense in the original parameter space $\mathcal{B}$ as the sample size grows, and

$$\hat{f}_{Y|X,Z^*}(y|x, z; B) = \frac{\sum_{i=1}^{n} K_{Y,h_Y}(Y_i - y) K_{X,h_X}(X_i - x) K_{Z,h_Z}(B(Z_i) - B(z))}{\sum_{i=1}^{n} K_{X,h_X}(X_i - x) K_{Z,h_Z}(B(Z_i) - B(z))}. \tag{30}$$

An asymptotic theory for this estimator is outside the scope of this work and is left for future research.

A semiparametric version is obtained by restricting $B(z)$ to be, e.g., linear so that

$$Z^* = BZ,$$

for some matrix $B \in \mathbb{R}^{d_Y \times d_Z}$. The corresponding estimator of $B(z)$ then takes the form $\hat{B}(z) = Bz$ where

$$\hat{B} = \arg \max_{\substack{B \in \mathbb{R}^{d_Y \times d_Z} \\ \|B\|=1}} \sum_{i=1}^{n} \log \hat{f}_{Y|X,Z^*}(Y_i | X_i, Z_i; B), \tag{31}$$

and $\hat{f}_{Y|X,Z^*}(Y_i|X_i, Z_i; B)$ takes the form in eq. (30) with $B(z) = Bz$. This semiparametric estimator was originally proposed in Fan et al (2009) as a dimension reduction device, and they show that $\hat{B}$ is $\sqrt{n}$-consistent. Thus, the first-step estimation of the index $\hat{Z}^* = \hat{B}Z$, will not affect the asymptotic properties of the final nonparametric estimators of $\bar{m}(x, z)$ as derived earlier.

The above estimator is still computationally challenging to implement when the number of parameters, $\dim(B) = d_Y d_Z$, is large since the numerical optimization problem in (31) cannot be solved in closed form. In this case, one can implement the following regression version of it: First, we choose as an approximate model the following partially linear regression, $Y = \mu(X) + \alpha BZ + u$, where $\mu : \mathbb{R}^{d_X} \mapsto \mathbb{R}^{d_Y}$ is a nonparametric regression function, while $\alpha \in \mathbb{R}^{d_Y \times d_Y}$ are the "factor loadings" of the index, BZ. Approximating μ by a linear combination of basis function, $\mu(x) \simeq \gamma p(x)$, where $\gamma \in \mathbb{R}^{d_Y \times K}$ are parameters and $p : \mathbb{R}^{d_X} \mapsto \mathbb{R}^K$ are basis functions, we obtain the following reduced rank regression,

$$Y = \gamma p(X) + \alpha BZ + u.$$

The estimates of γ, α and B can be computed using standard reduced rank OLS as originally proposed by Anderson (1951).

8 Counterfactual Demand Predictions with Consumer Scanner Data

Our methods are best applied to large data sets on individual behaviour. Here we examine consumer choices from a consumer home-scanner panel comprising more than 25,000 households whose behaviour is recorded electronically on a week by week basis over an eleven year period. The home-scan data, described in more detail below, collect granular records of expenditures, quantities and unit prices across a full range of groceries. The data also contain a large number of socio-economic variables (including family demographic structure, employment, education), behavioural measures (including smoking, newspaper readership, TV programmes), health (including BMI), and attitudinal data (including brand loyalty, attitudes to healthy living) which, we argue, are ideally suited to represent the 'external' covariates which play a key role in our approach to estimating individual level counterfactual demands.

After describing the data in the next subsection, we first use the data to estimate a parametric demand model to establish the average responses to price and income changes in the data. We then implement our approach to estimating individual counterfactual responses to price and income changes allowing for multidimensional unobserved heterogeneity. In our empirical analysis we find strong evidence that individual demands are nonlinear in relative prices and income, and that this nonlinear behaviour varies systematically across distinct individuals.[5]

8.1 The Homescan Data

The homescan data we use are from the Kantar Worldpanel and are collected via in-home scanning technology. They provide information on food categories that are purchased and brought into the home by a representative panel of British households over the eleven year period January 2005 to October 2016. There are approximately 26,000 households per year. Participants record spending on all grocery purchases via an electronic hand held scanner in the home. Purchases from all types of store (supermarkets, corner stores, online, local speciality shops) are covered by the data.

The data include information on the unit price paid for each product, detailed demographic details of the individuals in the household and the responses to attitudinal and behavioural questions, particularly on household shopper behaviour and media information. These data have been used in Dubois, Griffith and Nevo (2014) and Dubois, Griffith and O'Connell (2018), and similar data are widely used in the US, for example in Aguiar and Hurst (2007); see Griffith, Leibtag, Leicester and Nevo (2009) for a detailed comparison.

In our analysis we group goods into three aggregate categories:

1. *ambient goods* = bakery, drinks, cupboard ingredients, confectionery;
2. *fresh ingredients* = dairy, meat, fruit and vegetables;
3. *prepared foods* = chilled prepared.

Each subcategory comprises many (in most cases several 1000s of) products. For each subcategory we compute a mean price for each fascia-year-month (averaging over households and weeks within the month). We aggregate this to a fascia-time price series for each of the three aggregate goods by using constant expenditure weights across the whole sample for each of the subcategories.

As noted above the homescan data include a remarkably extensive set of individual characteristics and behavioural variables. We use these to form our external covariates Z that relate to individual preference heterogeneity. These variables variables are summarized in the following six broad categories:

1. *Main shopper and head of household characteristics*: Employment status; Marital status; Sex; Age; Date of birth; BMI.

2. *Household information*: Family type; TV Region; Acorn geodemographic code; Social class; No.

[5] In estimation we do not allow for the endogeneity of total expenditure or relative prices. In section 4.1, we show how the control function method could be applied to our new approach but we leave the empirical implementation of the estimator to future work.

of adults; No. of children; Household size; No. of females; No. in full time employment; Life-cycle€ stage of household; No. of smokers; No. of vegetarians; household member diabetic

3. *Household durable ownership information*: Household access to the internet; Cable television; Satellite television; Home computer; No. of cars; No. of televisions; Filter coffee maker; Dishwasher; Type of freezer; Microwave oven; Washing machine

4. *House tenure/type characteristics*: Housing tenure; No. of toilets; House has private garden; Type of building: Area type (e.g. town/village/city)

5. *Household shopping behaviour*: Freq. of main shop; Freq. of shopping by car; Main form of transport used to shop; Milk delivered

6. *Media information*: Main ITV channel; Other ITV channel; Days per week watching television; Hours per day watching television; Hours spent watching commercial stations.

Table 1: Consumer Data Descriptive Statistics

Discription		Mean	SD
Share(good 1): ambient	s_1	0.35	(0.17)
Share(good 2): fresh ingredients	s_2	0.49	(0.19)
Share(good 3): prepared	s_3	0.17	(0.13)
Log(total expenditure)	$\ln x$	1.76	(0.62)
Log(price good 1)	$\ln p_1$	-0.15	(0.20)
Log(price good 2)	$\ln p_2$	-0.37	(0.17)
First characteristics index	Z^*	17.77	(1.01)
Second characteristics index	Z^{**}	- 0.852	(1.75)
No. Obs.		260538	

Notes: Prices are relative to the numeraire good p_3. Total expenditure on the three agrregated goods x is deflated by share-weighted price index, using common shares across the whole sample. A full list of index characteristics (Z^* and Z^{**}) available on request.

Using the method described in section 7.2 we reduce these characteristics to two distinct index groups we label Z^* and Z^{**} described in Table 1.

8.2 Estimates from a Parametric Almost Ideal Model

To assess the overall properties of the consumer behaviour in this sample we begin by estimating a standard parametric Almost Ideal Model (Deaton and Muellbauer, 1980) with additive unobserved heterogeneity. For each consumer i, the budget share on good j is written

$$s_{ij} = \alpha_j + \sum_k \gamma_{jk} \ln p_k + \beta_j (\ln x_i - \ln P_i) + \epsilon_{ij}$$

where p_j is the price of good j, x is the total expenditure on all goods under consideration, ϵ_{ij} is unobservable individual and good specific heterogeneity, and where

$$\ln P = \alpha_0 + \sum_k (\alpha_k \ln p_{ik} + \sum_l \gamma_{lk} \ln p_{il} \ln p_{ik})).$$

Table 2: Almost Ideal Model Results: with unrestricted Z covariates

	(1) Share 1: ambient	(2) Share 2: fresh ingredients
β_j	-0.0202	0.0128
	(-34.11)	(19.87)
γ_{1k}	0.0862	-0.0530
	(34.31)	(-16.63)
γ_{2k}	-0.0626	0.0590
	(-23.00)	(15.60)
Observations	260538	260538

Notes: t-statistics in parentheses. Each share equation includes all Z covariates additively. The regressions use the Blundell and Robin (1999) algorithm for parametric Almost Ideal demand systems.

We allow the unobserved heterogeneity term ϵ_{ij} to be correlated across goods j, but assume mean independence with respect to the price and income variables. Since the model is linear in parameters, our estimates are consistent with a model with additive random heterogeneity in the parameters. A random coefficient model of this type would allow for multivariate unobserved heterogeneity but would not allow the identification of individual counterfactuals which motivates our approach.

Table 1 presents the estimated income and price parameters for the Almost Ideal specification. The estimated budget share model includes the complete set of Z covariates from the homescan data entered additively into each budget share equation. In these estimates linear homogeneity in prices and total expenditures is imposed so that there are two independent consumption choices within this set of three goods. The β_j parameter estimates suggest the aggregate of the ambient goods is a necessity while the fresh ingredient aggregate is a necessity. The γ_{jk} estimates suggest symmetry is not strongly at odds with the average responses in the data.

Table 3: Almost Ideal Model Results: with Z^* indices

	(1) Share 1: ambient	(2) Share 2: fresh ingredients
β_j	-0.0201	0.0168
	(-35.19)	(27.86)
γ_{1k}	0.0871	-0.0640
	(37.38)	(-19.93)
γ_{2k}	-0.0595	0.0600
	(-24.10)	(19.65)
Observations	260538	260538

Notes: t-statistics in parentheses. Each share equation include the two Z^* indices using the dimension reduction algorithm of section 7.2. The regressions follow Blundell and Robin (1999) algorithm for parametric Almost Ideal style demand systems.

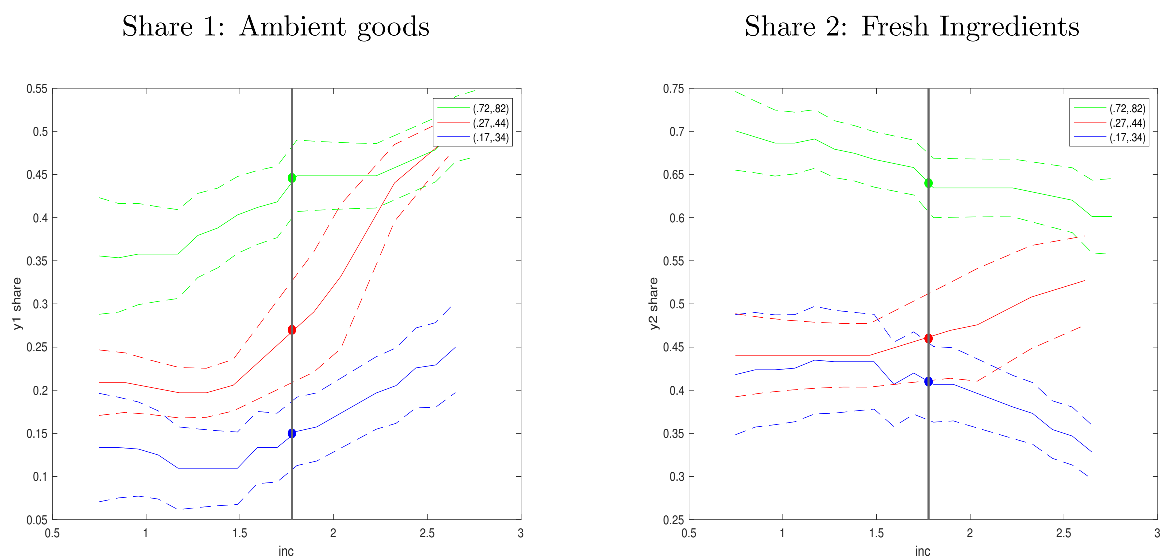
Table 2 presents the corresponding estimated parameters replacing the full set of covariates with the two indices Z^* calculated using the dimension reduction regression described in section 7.2. These results indicate that the reduction in dimension of the Z has little impact on the estimated price and income effects. In what follows we use these two indices to represent the excluded covariates throughout the analysis.

8.3 Estimated counterfactuals with unobserved heterogeneity

We now turn to the empirical implementation of our new approach. We present estimated price and income counterfactuals for individual demands allowing for multidimensional unobserved heterogeneity. There is no longer a single homogeneous price and income response for each price and income, but rather a distribution of estimated individual responses according to the distribution of unobserved heterogeneity.

In relation to the modelling framework in section 3, Y represents the vector of observed budget shares satisfying (1) where X contains the price and income variables, and ε represents the vector of unobserved heterogeneity. Recall that our objective is to estimate the responses in the budget shares (Y) to counterfactual changes in the income and price variables (X) for a distinct individual characterized by income and price vector $X = x$ and the initial vector of budget shares $Y = y$. Thus we estimate the counterfactual for the exogenous change of income or price x to x' for the *individual* described by $Y = y$.

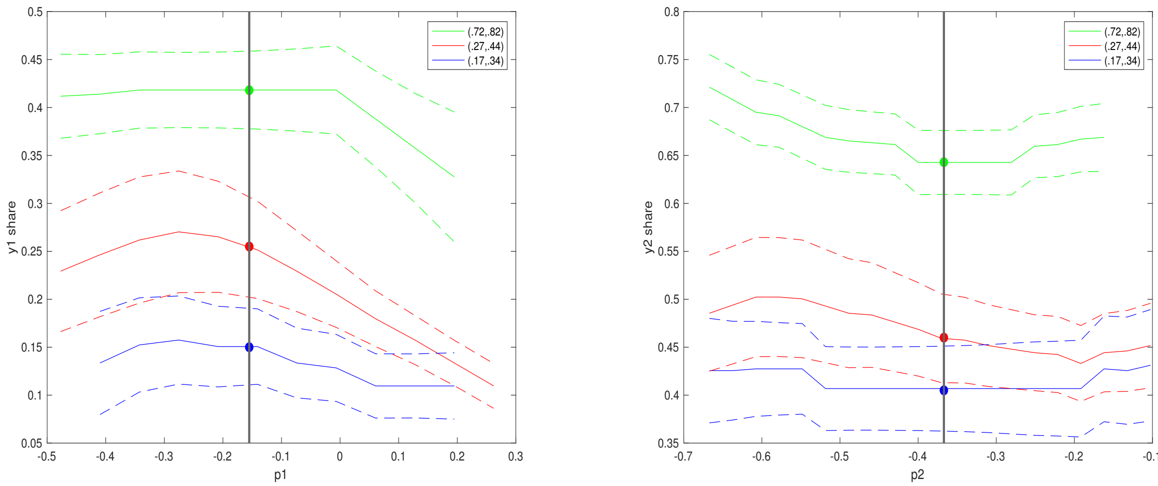
Figure 1: Estimated Individual Demand Responses by Income



Notes: The figures plot, with dots in the vertical line, the chosen shares of Ambient Foods and Fresh Ingredients at the median expenditures and median prices, of three households. The dots correspond to the quantiles, in brackets, of the marginal distributions of the two shares. The solid lines going through the dots are the estimated counterfactual demands for each of the three households, when varying total expenditure. The 95% confidence intervals are calculated from Theorem 11. Source: Kantar Homescan data.

Figure 2: Estimated Individual Demand Responses by by Price

Share 1: Ambient goods   Share 2: Fresh Ingredients



Notes: The figures plot, with dots in the vertical line, the chosen shares of Ambient Foods and Fresh Ingredients at the median expenditures and median prices, of three households. The dots correspond to the quantiles, in brackets, of the marginal distributions of the two shares. The solid lines going through the dots are the estimated counterfactual demands for each of the three households, when varying own price. The 95% confidence intervals are calculated from Theorem 11. Source: Kantar Homescan data.

In estimation we used a second order kernel. Each density and derivative of a density is estimated with bandwidths that vary depending on the number of arguments in the density and the order of the derivative. The bandwidth for the kth variable of (Y, X, Z) is chosen as

$$h_k = \ (std_k) \ (C) \ N^{-1/(4+I*d_Y+d_X+d_Z+4+2*j)}$$

where std_k is the standard deviation of the kth variable, C is a constant, $I = 1$ if the kth variable belongs to Y, and j is the number of derivatives. The grid algorithm by which the estimates were obtained is described in Appendix F.

In Figures 1 and 2 we plot the estimated counterfactual budget share responses of ambient goods (share 1) and fresh ingredients (share 2) to changes in total budget and relative prices, for different quantiles of the marginal distribution of the initial shares Y_1 and Y_2 respectively. These correspond to the three quartiles of the marginal distribution of Z^* and Z^{**}. The highlighted points on the vertical line in each figure are the observed data values for the three distinct individuals characterized by their observed shares. The estimated counterfactual demands for these three distinct individuals are then mapped out in the three continuous lines in each figure. Confidence intervals are constructed following the discussion in section 6. Note that, although there are three goods under study the impact on share 3 is given automatically since the sums of the shares equal unity.

Turning to the specific estimated counterfactuals, Figure 1 shows that the responses in the consumption of the two goods is nonlinear on total budget and that this nonlinearity varies strongly across individuals. For some types of individuals the plot of income responses declines with income while for other types it rises. Figure 2 indicates strong nonlinearities in the way individual consumers respond to prices. These responses and the shape of nonlinearity also vary systematically across individuals. Ignoring nonseparable multidimensional heterogeneity would miss these key features of behaviour.

9 Summary and Conclusions

We have developed methods for identifying and estimating counterfactuals in nonparametric models with nonseparable multidimensional unobserved heterogeneity. When values of the regressors change while the values of the unobserved variables stay fixed, our method allows to identify the new value of the dependent variables. We do not require any functional restrictions on the response function other than invertibitlity in the nonseparable vector of unobservable variables. The method is based on a transformation of variables equation and external variables that are used to define a mapping between derivatives of conditional densities of observable variables and derivatives of conditional densities of unobservable variables. Only when both derivatives equal zero, responses correspond to the same value of the unobservable variables. This relationship is used to trace responses when the values of the regressors change while those of the unobservable variables stay fixed. The regressors may be either discrete or continuously distributed.

We have extended the results in several directions, including the case where regressors and unobservable variables are not conditionally independent and cases where only partial identification is possible. We have also developed new identification results for identification of derivatives of structural functions in systems of simultaneous equations when the value of the unobservable variables stay fixed across different values of the regressors.

We have shown that our estimators are consistent and asymptotically normal, and we have applied them to estimate the heterogeneous responses of three households. The application used the UK Kantar homescan data. The households made different choices on a given budget set. Our method provided estimates for the choices these households would make if their tastes stayed the same while prices and/or total expenditures change. We found significant differences in the three households' estimated counterfactual choices, as well as significant differences with results obtained from using the data to estimate an Almost Ideal Demand model.

References

Aguiar, M. and E. Hurst (2007) Life-Cycle Prices and Production, American Economic Review, 97 (5): 1533-1559.

Ai, C. and X. Chen (2003), "Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1843.

Al-Sadoon, M.M. (2015) A General Theory of Rank Testing, Barcelona GSE Working Paper No 750.

Altonji, J. G. and Matzkin, R. L. (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," Econometrica, 73, 1053–1102.

Athey, S. and G. W. Imbens (2006) Identification and Inference in Nonlinear Difference-in-Differences Models, *Econometrica* 74, 431-497.

Beckert, W. (2006), "Specification and Identification of Stochastic Demand Models," Econometric Reviews, 26, 669–683.

Beckert, W. and R. Blundell (2008), "Heterogeneity and the Non-Parametric Analysis of Consumer Choice: Conditions for Invertibility," Review of Economic Studies, 75, 1069–1080.

Benkard, C.L. and S. Berry (2006) On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988), Econometrica 74, 1429-1440.

Beran, R., A. Feuerverger, & P. Hall (1996) On nonparametric estimation of intercept and slope distributions in random coefficient regression. Annals of Statistics 24, 2569-2692.

Beran, R. & P. Hall (1992) Estimating coefficient distributions in random coefficient regressions. Annals of Statistics 20, 1970-1984.

Beran, R. and P. W. Millar (1994), "Minimum Distance Estimation in Random Coefficient Regression Models," Annals of Statistics, 22, 1976–1992.

Berry, S., A. Gandhi and P. Haile (2013) Connected Substitutes and Invertibility of Demand, *Econometrica*, 81, 2087–2111.

Berry, S. and P. Haile (2009), "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," NBER Working Paper No. 15276.

Berry, S. and P. Haile (2014) Identification in Differentiated Products Markets Using Market Level Data, *Econometrica* 82, 1749–1797.

Berry, S. and P. Haile (2016) Identification in Differentiated Products Markets, Annual Review of Economics, 8, 27-52.

Berry, S. and P. Haile (2018) Identification of Nonparametric Simultaneous Equations Models with a Residual Index Structure, *Econometrica*, , 86, 289-315.

Berry, S., J. Levinsohn and A. Pakes (2004) Estimating Differentiated Product Demand Systems from a Combination of Micro and Macro Data: The New Car Model, *Journal of Political Economy* 112, 68–105.

Beckert, W. (2006) Specification and Identification of Stochastic Demand Models, Econometric Reviews, 26 (6), 669–683.

Beckert, W. and R. Blundell (2008) Heterogeneity and the Non-parametric Analysis of Consumer Choice: Conditions for Invertibility. *Review of Economic Studies*, 75, 1069-1080.

Blundell, R., X. Chen, and D. Kristensen (2007) Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves, *Econometrica*, 75, 1613-1669.

Blundell, R., J. Horowitz and M. Parey (2017) Nonparametric Estimation of a Nonseparable Demand Function under the Slutsky Inequality Restriction, *Review of Economics and Statistics* 99, 291–304.

Blundell, R., D. Kristensen, and R.L. Matzkin (2014) Bounding Quantile Demand Functions using Revealed Preference Inequalities, *Journal of Econometrics* 179, 112-127.

Blundell, R. and R. L. Matzkin (2014) Control Functions in Nonseparable Simultaneous Equations Models, *Quantitative Economics* 5, 271–295.

Brenier, Y. (1991) Polar Factorization and Monotone Rearrangement of VEctorvalued Functions. Communications on Pure and Applied Mathematics, 44, 375-417.

Briesch, R.A., P.K. Chintagunta, & R.L. Matzkin (2007) Nonparametric Discrete Choice Models with Unobserved Heterogeneity, Journal of Business & Economic Statistics, 28, 2, 291-307.

Brown, D. J., and R. Matzkin (1998) Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand, Discussion Paper 1175, Cowles Foundation

Brown, B.W. and M.B. Walker (1989) “The Random Utility Hypothesis and Inferences in Demand Systems”, Econometrica, 57, 815-829.

Carlier, G., V. Chernozhukov and A. Galichon (2016) Vector quantile regression: An optimal transport approach, *Annals of Statistics* 44, 1165-1192.

Chen, X. (2007) Large Sample Sieve Estimation of Semi-Nonparametric Models. *Handbook of Econometrics* 6, Part 2 (Eds. J.J. Heckman and E.E. Leamer), 5549–5632. Amsterdam, North-Holland.

Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014) Local Identification of Nonparametric and Semiparametric Models, *Econometrica*, , 82, 785-809.

Chen, X. and D. Pouzo (2009) Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals, *Econometrica*, , 80, 277-321.

Chernozhukov, V., A. Galichon, M. Henry, and B. Pass (2020) Identification of Hedonic Equilibrium and Nonseparable Simultaneous Equations. Journal of Political Economy, forthcoming.

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

Chernozhukov, V., G. Imbens, and W. Newey (2007) Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions, *Journal of Econometrics.*

Chesher, A. (2003), “Identification in Nonseparable Models,” Econometrica, 71, 1405–1441.

Chiappori, P.A., D. Kristensen and I. Komunjer (2016a) On Nonparametric Identification of Multiple Choice Models, unpublished manuscript, UCL.

Chiappori, P.A., D. Kristensen and I. Komunjer (2016b) Nonparametric Identification in Nonlinear Simultaneous Equations Models: The Case of Covariance Restrictions, unpublished manuscript, UCL.

Cuhna, F., J. J. Heckman, and S. Shennach (2010) Estimating the Technology of the Cognitive and Noncognitive Skill Formation, *Econometrica* 78, 883–931.

D’Haultfoeuille and Fevrier (2015) Identification of Nonseparable Triangular Models With Discrete Instruments, *Econometrica,* Vol. 83, No. 3, 1199-1210.

Darolles. S., Y. Fan, J. P. Florens, and E. Renault (2011) Nonparametric Instrumental Regression, *Econometrica,* 79, 1541-1565.

Deaton, A. and J. Muellbauer (1980) Economics and Consumer Behavior, Cambridge University Press, Cambridge.

Dubois, Griffith and O'Connell (2018) "The effects of banning advertising in junk food markets" Review of Economic Studies 85, 1, 396-436

Dubois, Griffith and Nevo (2014) "Do Prices and Attributes Explain International Differences in Food Purchases?" American Economic Review, 104, 3, 832-867

Ekeland, I., A. Galichon, and M. Henry (2012) Comonotone Measures of Multivariate Risks, Mathematical Finance, 22, 109-132.

Feuerverger, A. & Y. Vardi (2000) Positron emission tomography and random coefficients regression. Annals of the Institute of Statistical Mathematics 52, 123-138.

Foster, A. & J. Hahn (2000) A consistent semiparametric estimation of the consumer surplus distribution. Economics Letters 69, 245-251.

Fox, J.T., K. I. Kim, S. P. Ryan, and P. Bajari (2012), "The random coefficients logit model is identified," Journal of Econometrics, 166, 204-212.

Fox, J.T. and A. Gandhi, (2011) Identifying Demand with Multidimensional Unobservables: A Random Functions Approach, Working Paper, University of Michigan, 2011.

Fox, J.T. and A. Gandhi, (2016) Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models, RAND Journal of Economics, 47, 1, 118-139.

Gale, D., and H. Nikaido (1965) The Jacobian Matrix and Global Univalence of Mappings, *Mathematische Annalen* 159, 81–93.

Galichon, A. and M. Henry (2012) Dual Theorey of Choice under Multivariate Risks, Journal of Economic Theory, 147, 1501-1516.

Gautier, E. and Y. Kitamura (2013), Nonparametric Estimation in Random Coefficients Binary Choice Models, Econometrica, 81, 2, 581-607.

Griffith, Leibtag, Leicester and Nevo (2009) "Consumer shopping behavior: how much do consumers save?" Journal of Economic Perspectives, 23,2, 99-120.

Hall P. and J.L. Horowitz (2005) Nonparametric Methods for Inference in the Presence of Instrumental Variables, Annals of Statistics, Vol. 33, 2904-2929.

Hausman, J. and D. Wise (1978) A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences, Econometrica, vol. 46, 2, 403-426.

Heckman, J.J. (1974) Effects of Day-Care Programs on Women's Work Effort, Journal of Political Economy, 782.

Heckman, J.J. and R. Willis (1977) A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women, Journal of Political Economy.

Hildreth, C. & J.P. Huock (1968) Some estimators for a linear model with random coefficients. Journal of the American Statistical Association 63, 584-92.

Hoderlein, S., H. Holzmann, and A. Meister (2017), The Triangular Model with Random Coefficients, Journal of Econometrics, vol. 201, 1, 144-169.

Hoderlein, S., Klemelae, J. and Mammen, E. (2010), "Analyzing the Random Coefficient Model Nonparametrically," Econometric Theory, 26, 804-837.

Hoderlein, S. and Mammen, E. (2007), "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," Econometrica, 75, 1513–1518.

Hoderlein, S. and S. Mihaleva (2008), "Increasing the price variation in a repeated cross section," Journal of Econometrics, 147, 316-25.

Hoderlein, S, L. Nesheim and A. Simoni, (2016), Semiparametric Estimation of Random Coefficients in Structural Economic Models, Econometric Theory.

Hoderlein, S., and J. Stoye (2014) Revealed Preferences in a Heterogeneous Population, *Review of Economics and Statistics* 96, 197-213.

Hoderlein, S., and J. Stoye (2015) Testing Stochastic Rationality and Predicting Stochastic Demand: The Case of Two Goods, *Economic Theory Bulletin,* 3, 313–328.

Hoderlein, S. and A. Vanhems (2017), "Welfare Analysis using Nonseparable Models," *Journal of Applied Econometrics.*

Horowitz, J.L. and S. Lee (2007) Nonparametric Instrumental Variables Estimation of a Quantile Regression Model, *Econometrica* 75, 1191-1208.

Hurwicz, L. (1950) Systems with nonadditive disturbances. In: Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph, vol.10. JohnWiley, NewYork

Ichimura, H. Thompson, T.S. (1998), Maximum Likelihood estimation of a binary choice model with random coefficients of unknown distribution, J. of Econometrics 86, 269-295.

Imbens, G. and W.K. Newey (2009) Identification and Estimation of Triangular Simultaneous Equations Models without Additivity, *Econometrica* 77, 1481-1512.

Kitamura, Y. and J. Stoye (2018) Nonparametric Analysis of Random Utility Models, Econometrica, 86, 6, 1883-1909.

Lancaster, T. (1979) Econometric Methods for the Analysis of Unemployment, Econometrica, 47, 939-956.

Lehman, E.L., 1974. Nonparametrics: Statistical Methods Based on Ranks. Holden-Day, San Francisco.

Lewbel, A. (2001), "Demand Systems with and without Errors," American Economic Review, 91, 611-618.

Lewbel, A. and K. Pendakur (2017) Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients, *Journal of Political Economy* , 125, 1100–1148.

Masten, M. (2015) Random Coefficients on Endogenous Variables in Simultaneous Equations Models, Review of Economic Studies, 85, 2, 1193-1250.

Matzkin, R.L. (2003) Nonparametric estimation of nonadditive random functions, *Econometrica* 71, 1339-1375.

Matzkin, R.L. (2007) Heterogeneous Choice, in *Advances in Econometrics: Proceedings of the 9th World Congress*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

Matzkin, R. L. (2007a), Nonparametric Identification, in: J. Heckman and E. Leamer, Eds, Handbook of Econometrics, vol. 6B, chap. 73, 5307-5368, New York: Elsevier Science.

Matzkin, R. L. (2007b), "Heterogeneous Choice," in R. Blundell, W. Newey and T. Pers-son, eds., Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, vol. 3, ch. 4, Econometric Society Monographs, 43 Cambridge: Cambridge University Press, 75–110.

Matzkin, R.L. (2008) Identification in Nonparametric Simultaneous Equations Models, *Econometrica* 76, 945–978.

Matzkin, R.L. (2012) Identification of Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity, *Journal of Econometrics*, 166, 106–115.

Matzkin, R.L. (2013) Nonparametric Identification in Structural Economic Models, Annual Review of Economics, 5, 547-486.

Matzkin, R.L. (2015) Estimation of Nonparametric Models with Simultaneity, *Econometrica* 83, 1-66.

McCann, R. J. (1995) Existence and uniqueness of monotone measure-preserving maps, Duke Math. J. 80 309–323.

McElroy, M.B. (1987) Additive General Error Models for Production, Cost, and Derived Demand or Share Systems, *Journal of Political Economy*, 95(4), 737-757.

McFadden, D. (1974) Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka, ed. Frontiers in Econometrics, pp105-142, New York: Academic Press.

McFadden, D.L. (2005) Revealed Stochastic Preference: A Synthesis, *Economic Theory*, 26, 245–264.

McFadden, D. and M. K. Richter (1991), "Stochastic Rationality and Revealed Stochastic Preference," in J. Chipman, D. McFadden, and M.K. Richter (eds.), Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, 161-186, Westview Press: Boulder, CO.

Newey, W. and D. McFadden (1994) Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics* 4 (Eds. R. Engle and D. McFadden), 2113-2245. Amsterdam, North-Holland.

Newey, W. and J. Powell (2003) "Instrumental Variables Estimation of Nonparametric Models," *Econometrica.*

Swamy, P.A.V.B. (1970) Efficient inference in a random coefficient model. Econometrica 38(2), 311-23.

Torgovitsky (2015) Identification of Nonseparable Models Using Instruments With Small Support, *Econometrica, Vol. 83, No.3, 1185-1197.*

Villani, C. (2003). Topics in Optimal Transportation. Graduate Studies in Mathematics. Amer. Math. Soc., Providence, RI.

Appendices

A Proofs

Proof of Theorem 2. The result follows by noticing that Assumption 2' guarantees that when the value of Z stays fixed, the value of ε stays fixed as well at $e = \Lambda(z)$. Assumptions 1 and 3 are then used to find the values of Y corresponding to $m(X, e)$. Specifically, let z be given and let $e = \Lambda(z)$ denote the unique value of ε whose existence is guaranteed by Assumption 2'. By invertibility (Assumption 1), for any value $x \in \mathcal{X}$, the value of Y satisfying $y = m(x, e)$ must also satisfy $e = r(y, x)$. By the definition of Λ, for $e = \Lambda(z)$,

$$\frac{\partial \log f_{\varepsilon|Z}(r(y,x)|z)}{\partial z} = \frac{\partial \log f_{\varepsilon|Z}(\Lambda(z)|z)}{\partial z} = 0$$

By (relationship), the value y of Y satisfying $y = m(x, e)$ satisfies

$$\frac{\partial \log f_{Y|X,Z}(y|x,z)}{\partial z} = \frac{\partial \log f_{\varepsilon|Z}(r(y,x)|z)}{\partial z}$$

Hence, for any $x, x' \in \mathcal{X}$ and $e = \Lambda(z)$, the values $y, y' \in \mathcal{Y}$ such that $y = m(x, e)$ and $y' = m(x', e)$ are those satisfying

$$\frac{\partial \log f_{Y|X,Z}(y'|x',z)}{\partial z} = 0 \qquad \text{and} \qquad \frac{\partial \log f_{Y|X,Z}(y'|x',z)}{\partial z} = 0$$

It follows that $\Delta_e(x, x') = m(x', e) - m(x, e)$ is identified. ■

Proof of Theorem 3. As in the proof of Theorem 2, Assumptions 1,2', and 3 imply that for any $x \in \mathcal{X}$ and $z \in \mathcal{Z}$, the value of the function $m(x, \Lambda(z))$ is identified as the value y satisfying the equation

$$\frac{\partial \log f_{Y|X,Z}(y|x,z)}{\partial z} = 0$$

Let $x \in \mathcal{X}$ and $e \in \mathcal{E}$ be given. By Assumption 4, there exists $z(e) \in \mathcal{Z}$ such that $\Lambda(z) = e$. Since Λ is known, $z(e)$ is known. The value $m(x, e)$ is then the value y satisfying the equation

$$\frac{\partial \log f_{Y|X,Z}(y|x,z(e))}{\partial z} = 0$$

■

Proof of Theorem 4. Let $\bar{\pi}^{-1}(w, x_1)$ denote the inverse of $\pi(w, T(t))$ w.r.t. t which exists and is identified under Assumption 3*(i)-(ii). Thus, we can treat $\bar{\eta} = T^{-1}(\eta) = \bar{\pi}^{-1}(W, X_1)$ as observed. Since T is one-to-one, the sigma algebra associated with η is equal to that associated with $\bar{\eta}$ so that conditional distributions given η are identical to those given $\bar{\eta}$.

We now show that X and ε are mutually independent conditional on $(Z, \bar{\eta})$ so that $\varepsilon|\left(X, Z, \bar{\eta}\right) \stackrel{d}{=} \varepsilon|\left(Z, \bar{\eta}\right)$. First, for any bounded function $a(X)$, by conditional independence of W and (ε, η) under Assumption 3*(iii),

$$E[a(X)|\varepsilon, \bar{\eta}, Z] = E[a(\pi\left(W, \eta\right))|\varepsilon, \eta, Z] = \int a(\pi\left(w, \eta\right))dF_{W|Z}\left(w|Z\right) = E[a(X)|\eta, Z],$$

which in turn implies that, for any bounded function $b\left(\varepsilon\right)$,

$$\begin{aligned} E[a(X)b\left(\varepsilon\right)|\bar{\eta}, Z] &= E\left[E\left[a(X)|\varepsilon, \bar{\eta}, Z\right]b\left(\varepsilon\right)|\bar{\eta}, Z\right] \\ &= E\left[E\left[a(X)|\bar{\eta}, Z\right]b\left(\varepsilon\right)|\bar{\eta}, Z\right] \\ &= E\left[a(X)|\bar{\eta}, Z\right]E\left[b\left(\varepsilon\right)|\bar{\eta}, Z\right] \end{aligned}$$

as desired.

Finally, observe that the results of Theorems all are based on the identity stated in eq. (6). We now show that a similar identity holds once we include $\bar{\eta}$ as additional control variable: First, observe that, due to the conditional independence betweeen X and ε as shown above,

$$f_{Y|X,Z,\bar{\eta}}(y|x, z, t) = f_{\varepsilon|Z,\bar{\eta}}\left(r\left(y, x\right)|z, t\right)\left|\frac{\partial r\left(y, x\right)}{\partial y}\right|,$$

where $f_{\varepsilon|Z,\bar{\eta}}\left(e|z, t\right)$ is well-defined due to Assumption 3*(iv). This in turn implies that

$$\frac{\partial \log f_{Y|X,Z,\bar{\eta}}(y|x, z, t)}{\partial z} = \frac{\partial \log f_{\varepsilon|Z,\bar{\eta}}\left(r\left(y, x\right)|z, t\right)}{\partial z}.$$

All arguments used in the proofs of Theorems 1–3 now go through for any fixed value of t. ■

Proof of Theorem 5. Assumptions 1 and 3 and equation (7) imply that for all $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $z \in \mathcal{Z}$ and $e = r\left(y, x\right) \in \mathcal{E}$

$$\frac{\partial \log f_{Y|X,Z}(y|x, z)}{\partial z_j} = \frac{\partial \log f_{\varepsilon|Z}\left(r\left(y, x\right)|z\right)}{\partial z_j}$$

Differentiation with respect to y, yields

$$\frac{\partial \log f_{Y|X,Z}(y|x, z)}{\partial z_j \partial y'} = \frac{\partial \log f_{\varepsilon|Z}\left(r\left(y, x\right)|z\right)}{\partial z_j \partial \varepsilon'}\left(\frac{\partial r\left(y, x\right)}{\partial y}\right) \tag{32}$$

By Assumption 5, when $z = z^*$ is such that

$$\frac{\partial \log f_{Y|X,Z}(y|x, z^*)}{\partial z} = 0$$

the coordinates of the vector $\partial \log f_{\varepsilon|Z}\left(r\left(y,x\right)|z\right)/\partial z_j \partial \varepsilon'$ are all zero except for the $j-th$ coordinate. Denote the value of such coordinate by a_j. It follows then from (32) that

$$\frac{\partial \log f_{Y|X,Z}(y|x,z^*)}{\partial z_j \partial y'} = a_j \ \left(\frac{\partial r^j\left(y,x\right)}{\partial y'}\right). \tag{33}$$

The function $s^j\left(y_{-j},x,e_j\right)$ can be defined implicitly by substituting y_j with $s^j\left(y_{-j},x,e_j\right)$ in the equation

$$r^j\left(y,x\right) = e_j$$

Hence, the Implicit Function Theorem implies that

$$\frac{\partial s^j\left(y_{-j},x,e_j\right)}{\partial y'_{-j}} = -\left(\frac{1}{\partial r^j\left(y,x,e_j\right)/\partial y_j}\right) \ \frac{\partial r^j\left(y,x,e_j\right)}{\partial y'_{-j}}$$

It follows by (33) that $\partial s^j\left(y_{-j},x,e_j\right)/\partial y'_{-j}$ is identified by the equation

$$\frac{\partial s^j\left(y_{-j},x,e_j\right)}{\partial y'_{-j}} = -\left(\frac{\partial \log f_{Y|X,Z}(y|x,z^*)}{\partial z_j \partial y'_j}\right)^{-1} \ \left(\frac{\partial \log f_{Y|X,Z}(y|x,z^*)}{\partial z_j \partial y'_{-j}}\right)$$

To show that $\partial s^j\left(y'_{-j},x',e_j\right)/\partial y'_{-j}$ is identified, we note that Assumption 5 applies also to $r\left(y',x'\right)$ because by the definition of y'

$$r\left(y',x'\right) = r\left(y,x\right)$$

Hence, by the same above arguments as for $\left(y,x\right)$, it follows that

$$\frac{\partial s^j\left(y'_{-j},x',e_j\right)}{\partial y'_{-j}} = -\left(\frac{\partial \log f_{Y|X,Z}(y'|x',z^*)}{\partial z_j \partial y'_j}\right)^{-1} \ \left(\frac{\partial \log f_{Y|X,Z}(y'|x',z^*)}{\partial z_j \partial y'_{-j}}\right)$$

This completes the proof. ■

Proof of Theorem 8. Since U_1 and U_2 are strictly monotone in $\left(y, y_{d_Y+1}\right)$ and y, respectively, the budget constraint is satisfied with equality. Moreover, since for each ε, U is strictly quasiconcave, the value of y that solves the first order conditions for the maximization of U when $y_{d_Y+1} = I - p'y$ is unique. Let $s\left(y,p,I,\varepsilon\right)$ denote the vector of the d_Y functions such that $s\left(y,p,I,\varepsilon\right) = 0$ denotes this system of first order conditions, and let $y = m\left(p,I,\varepsilon\right)$ denote the demand function which satisfies $s\left(m\left(p,I,\varepsilon\right),p,I,\varepsilon\right) = 0$. We will show that for each $\left(y,p,I\right)$, $\varepsilon \mapsto s\left(y,p,I,\varepsilon\right)$ is globally univalent (see Gale and Nikaido, 1965). This guarantees the global existence of an implicit function $r\left(y,p,I\right)$ such that for all $\left(y,p,I\right)$ in a region, $s\left(y,p,I,r\left(y,p,I\right)\right) = 0$. The uniqueness of m on ε and of r on y, for any $\left(p,I\right)$, imply that

$$y = m\left(p,I,\varepsilon\right) \Leftrightarrow \varepsilon = r\left(y,p,I\right).$$

Hence, the demand function $m(p, I, \varepsilon)$ is invertible in ε.

To show that $s(y, p, I, \cdot)$ is globally univalent in ε, we note that for each y and with $y_{d_Y+1} = I - p'y$,

$$s(y, p, I, \varepsilon) = \begin{bmatrix} \frac{\partial U_1}{\partial y_1} - p_1 \frac{\partial U_1}{\partial y_{d_Y+1}} + \frac{\partial U_2}{\partial y_1} \\ \vdots \\ \frac{\partial U_1}{\partial y_{d_Y}} - p_{d_Y} \frac{\partial U_1}{\partial y_{d_Y+1}} + \frac{\partial U_2}{\partial y_{d_Y}} \end{bmatrix}.$$

Since only U_2 is a function of ε, the Jacobian of $s(y, p, I, \varepsilon)$ with respect to ε equals $D_{y\varepsilon} U_2(y, \varepsilon)$ as defined in the theorem. The assumption on the determinant of the principal minors imply that $D_{y\varepsilon} U_2(y, \varepsilon)$ is a so-called P-matrix, and so it follows by Gale and Nikaido (1965) that $\varepsilon \mapsto s(y, p, I, \varepsilon)$ is globally univalent. ■

Proof of Theorem 9. First note that under Assumptions 1-5, the density of (Y, X, Z), $f_{Y,X,Z}(y, x, z)$, is twice continuously differentiable. Thus, employing standard results for kernel density estimation, it holds that, for any given $(x, z) \in X_0 \times Z$,

$$\sup_{y \in \mathcal{Y}_0} \|\hat{g}(y|x, z) - g(y|x, z)\| = O_P\left(h_Y^2\right) + O_P\left(h_X^2\right) + O_P\left(h_Z^2\right) + O_P\left(\frac{\log(n)}{n h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}}\right),$$

where $\hat{g}(y|x, z)$ is defined in eq. (20), and $g(y|x, z) = \partial f_{Y|X,Z}(y|x, z)/(\partial z)$; see, for example, Hansen (2008, Theorem 7). This combined with Assumption 6(i) and the bandwidth restrictions stated in the theorem yield

$$\left\|\hat{g}(y|x, z)' \hat{g}(y|x, z) - g(y|x, z) g(y|x, z)\right\| = o_P(1).$$

Consistency now follows from Newey and McFadden (1994, Theorem 2.6), where identification is achieved through Theorem 1.

Next, we derive the asymptotic distribution of $\hat{m}(x, z)$: With $\hat{y}^* := \hat{m}(x, z)$, $y^* := \bar{m}(x, z)$ and $\tilde{y}$ situated on the line segment connecting $\hat{y}^*$ and y^*, the first-order condition for $\hat{y}^*$ together with the mean value theorem yield

$$0 = \hat{G}(x, z)' \hat{g}(y^*|x, z) + \hat{H}_m(x, z)(\hat{y}^* - y^*),$$

where $\hat{G}(x, z) = \partial^2 \hat{f}_{Y|X,Z}(y^*|x, z)/(\partial z \partial y')$, $\hat{H}_m(x, z) := \hat{G}(x, z)' \tilde{G}(x, z)$, and $\tilde{G}(x, z) = \partial^2 \hat{f}_{Y|X,Z}(\tilde{y}|x, z)/(\partial z \partial y')$. Under the stated conditions on the bandwidths in eq. (25), it follows from Lemma 13 that

$$\sqrt{n h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}} \{\hat{g}(y^*|x, z) - g(y^*|x, z)\} = \sqrt{n h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}} \hat{g}(y^*|x, z) \to^d N(0, \Omega_m(x, z)), \quad (34)$$

while $\hat{G}(x, z)$ and $\tilde{G}(x, z)$ both converge towards $G(x, z)$ in probability. The claimed asymptotic distribution result now follows by the same arguments as in the proof of Newey and McFadden (1994, Theorem 2.6). ■

Proof of Theorem 10. The proof of the theorem proceeds along the same lines as the one for Theorem 9, and so we only sketch the proof for $\hat{r}(y,x)$. With $\hat{g}(y|x,z)$ defined in eq. (20) and $g(y|x,z) = \partial f_{Y|X,Z}(y|x,z)/(\partial z)$, we have, for any given $(y,x) \in X \times Y$,

$$\sup_{z\in\bar{\mathcal{E}}_0} \|\hat{g}(y|x,z) - g(y|x,z)\| = O_P\left(h_Y^2\right) + O_P\left(h_X^2\right) + O_P\left(h_Z^2\right) + O_P\left(\frac{\log(n)}{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2}}\right).$$

This together with the identification result of Theorem 1 shows consistency. To obtain the asymptotic distribution, first observe that, with $\hat{e}^* := \hat{r}(y,x)$ and $z^* := r(y,x)$,

$$0 = \hat{g}(y|x,\hat{e}^*) = \hat{g}(y|x,z^*) + \hat{H}_r(y,x)(\hat{e}^* - z^*),$$

where $\hat{H}_r(y,x) = \partial^2 \hat{f}_{Y|X,Z}(y|x,\tilde{e})/(\partial z \partial z')$, and $\tilde{e}$ is situated on the line segment connecting $\hat{e}^*$ and z^*. Under the bandwidth conditions, Lemma 13 implies that

$$\sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2}}\left\{\hat{g}(y|x,z^*) - g(y|x,z^*)\right\} = \sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2}}\hat{g}(y|x,z^*) \to^d N(0,\Omega_r(y,x)), \quad (35)$$

and $\hat{H}_r(y,x) \to^P H_r(y,x)$, where $\Omega_r(y,x) \in R^{d_Y\times d_Y}$ and $H_r(y,x) \in R^{d_Y\times d_Y}$ are defined in the theorem. ■

Proof of Theorem 11. Write

$$\hat{\Delta}_{\Lambda(z)}\left(x,x'\right) - \Delta_{\Lambda(z)}\left(x,x'\right) = \{\hat{m}\left(x',z\right) - \bar{m}\left(x',z\right)\} - \{\hat{m}(x,z) - \bar{m}(x,z)\}, \quad (36)$$

where the two terms on the right-hand side are asymptotically independent of each other by the usual arguments for kernel-based estimators. Thus, $\hat{\Delta}_{z_0}(x,x')$'s large-sample distribution follows directly from Theorem 9.

Similarly,

$$\begin{aligned}\hat{\Delta}_y\left(x,x'\right) - \tilde{\Delta}_y\left(x,x'\right) &= \hat{m}(x',\hat{r}(y,x)) - \bar{m}(x',\bar{r}(x,y)) \\ &= \frac{\partial \hat{m}(x',\bar{z})}{\partial z'}\{\hat{r}(y,x) - \bar{r}(y,x)\} + \left\{\hat{m}(x',\bar{r}(x,y)) - m(x',\bar{r}(x,y))\right\},\end{aligned}$$

where $\bar{z}$ is situated on the line segment connecting $\hat{r}(y,x)$ and $\bar{r}(x,y)$; in particular, $\bar{z} \to^P \bar{r}(x,y)$ and so $\partial\hat{m}(x',\bar{z})/(\partial z') \to^P \partial\bar{m}(x',\bar{r}(x,y))/(\partial z')$. Since $\hat{r}(y,x)$ and $\hat{m}(x',\bar{r}(x,y))$ are independent in large samples by the usual arguments for kernel estimators, the large sample distribution now follows by combining Theorems 9 and 10. ■

Proof of Theorem 12. It is easily checked that all the arguments in the proofs of Theorems 9–11 remain true with $\hat{f}_{Y|X,Z}(y|x,z)$ on the form (28), except that we can no longer employ Lemma 13 to derive the limiting distribution of $\hat{g}(y|x,z) = \partial\hat{f}_{Y|X,Z}(y|x,z)/(\partial z)$. Inspecting the proof of

Lemma 13, we see that all arguments remain unchancged except for limiting distribution as stated in (39) since this assumes no time series dependence, which is present in the panel data case. To derive the limiting distribution under panel data time series dependence, first note that in the panel data case, $\partial \hat{f}_{Y,X,Z}(y,x,z)/(\partial z) = \sum_{t=1}^{T} \partial \hat{f}^{(t)}_{Y|X,Z}(y,x,z)/(\partial z)/T$ where

$$\hat{f}^{(t)}_{Y|X,Z}(y,x,z) = \sum_{i=1}^{n} K_{Y,h_Y}(Y_{i,t}-y) K_{X,h_X}(X_{i,t}-x) K_{Z,h_Z}(Z_{i,t}-z). \tag{37}$$

Second, by standard arguments for kernel density estimators with i.i.d. data, we find that that $D\hat{F}(y,x,z) := \left(\partial \hat{f}^{(1)}_{Y|X,Z}(y,x,z)/(\partial z'), ...\partial \hat{f}^{(T)}_{Y|X,Z}(y,x,z)/(\partial z')\right)'$ satisfies

$$\sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}} \left\{ D\hat{F}(y,x,z) - DF(y,x,z) - \sum_{a\in\{Y,X,Z\}} h_a^2 B_a(y|x,z) \right\} \to^d N\left(0, \tilde{V}_F(y,x,z)\right), \tag{38}$$

where $\tilde{V}_F(y,x,z) = \left[\tilde{V}_{F,st}(y,x,z)\right]_{s,t=1}^{T}$ with $\tilde{V}_{F,tt}(y,x,z) = \tilde{V}(y,x,z)$, with $\tilde{V}(y,x,z)$ is defined in eq. (40), while, for $s \neq t$,

$$\tilde{V}_{F,st}(y,x,z) = \lim_{n\to\infty} nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} \mathrm{Cov}\left(\frac{\partial \hat{f}^{(s)}_{Y|X,Z}(y,x,z)}{\partial z'}, \frac{\partial \hat{f}^{(t)}_{Y|X,Z}(y,x,z)}{\partial z'} \right).$$

What remains is to derive $\tilde{V}_{F,st}(y,x,z)$. To this end, note that

$$\begin{aligned}
& nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} \mathrm{Cov}\left(\frac{\partial \hat{f}^{(s)}_{Y|X,Z}(y,x,z)}{\partial z'}, \frac{\partial \hat{f}^{(t)}_{Y|X,Z}(y,x,z)}{\partial z'} \right) \\
= \ & nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} E\left[\left(\frac{1}{n} \sum_{i=1}^{n} K_{Y,h_Y}(Y_{i,s}-y) K_{X,h_X}(X_{i,s}-x) K^{(1)}_{Z,h_Z}(Z_{i,s}-z) \right) \right. \\
& \left. \times \left(\frac{1}{n} \sum_{j=1}^{n} K_{Y,h_Y}(Y_{j,t}-y) K_{X,h_X}(X_{j,t}-x) K^{(1)}_{Z,h_Z}(Z_{j,t}-z) \right) \right] + o(1) \\
= \ & h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} E\left[\sum_{i=1}^{n} K_{Y,h_Y}(Y_{i,s}-y) K_{X,h_X}(X_{i,s}-x) K^{(1)}_{Z,h_Z}(Z_{i,s}-z) \right. \\
& \left. \times K_{Y,h_Y}(Y_{i,t}-y) K_{X,h_X}(X_{i,t}-x) K^{(1)}_{Z,h_Z}(Z_{i,t}-z) \right] + o(1) \\
= \ & h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} \int f_{(Y_s,X_s,Z_s),(Y_t,X_t,Z_t)}(y_s,x_s,z_s,y_t,x_t,z_t) K_{Y,h_Y}(y_s-y) K_{X,h_X}(x_s-x) K^{(1)}_{Z,h_Z}(z_s-z) \\
& \times K_{Y,h_Y}(y_t-y) K_{X,h_X}(x_t-x) K^{(1)}_{Z,h_Z}(z_t-z)\, d(y_s,x_s,z_s,y_t,x_t,z_t) \\
= \ & h_Y^{d_Y} h_X^{d_X} h_Z^{d_Z} f_{(Y_s,X_s,Z_s),(Y_t,X_t,Z_t)}(y,x,z,y,x,z) \left(\int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z(z)}{\partial z} dz \right)^2 \\
= \ & 0,
\end{aligned}$$

where we have used that $\int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z(z)}{\partial z} dz = 0$. ■

B Lemmas

Lemma 13 *Suppose that Assumptions A.1-A.5 hold. Then:*

1. *As* $nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}h_a^4 \to 0$ *for* $a \in \{Y, X, Z\}$, *and* $nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2} \to \infty$,

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}}\left\{\frac{\partial \hat{f}_{Y|X,Z}(y|x,z)}{\partial z} - \frac{\partial f_{Y|X,Z}(y|x,z)}{\partial z}\right\} \to^d N\left(0, V\left(y,x,z\right)\right),$$

where

$$V\left(y,x,z\right) = \frac{f_{Y|X,Z}(y|x,z)}{f_{X,Z}\left(x,z\right)}\int_{\mathbb{R}^{d_Y}} K_Y^2\left(y\right)dy\int_{\mathbb{R}^{d_X}} K_X^2\left(x\right)dx\int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z\left(z\right)}{\partial z}\frac{\partial K_Z\left(z\right)}{\partial z'}dz \in \mathbb{R}^{d_Z \times d_Z}.$$

2. *As* $h_X, h_X \to 0$ *and* $nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+4} \to \infty$,

$$\frac{\partial^2 \hat{f}_{Y|X,Z}(y|x,z)}{\partial z \partial z'} \to^P \frac{\partial^2 f_{Y|X,Z}(y|x,z)}{\partial z \partial z'}.$$

3. *As* $h_X, h_X \to 0$ *and* $nh_Y^{d_Y+2}h_X^{d_X}h_Z^{d_Z+2} \to \infty$,

$$\frac{\partial^2 \hat{f}_{Y|X,Z}(y|x,z)}{\partial z \partial y'} \to^P \frac{\partial^2 f_{Y|X,Z}(y|x,z)}{\partial z \partial y'},$$

Proof. We have

$$\frac{\partial \hat{f}_{Y|X,Z}(y|x,z)}{\partial z} = \hat{f}_{X,Z}^{-1}(x,z)\frac{\partial \hat{f}_{Y,X,Z}(y,x,z)}{\partial z} + \frac{\hat{f}_{Y,X,Z}(y|x,z)}{\hat{f}_{X,Z}^2(x,z)}\frac{\partial \hat{f}_{X,Z}(x,z)}{\partial z},$$

where

$$\begin{aligned}
\hat{f}_{Y,X,Z}(y,x,z) &= \sum_{i=1}^{n} K_{Y,h_Y}\left(Y_i - y\right)K_{X,h_X}\left(X_i - x\right)K_{Z,h_Z}\left(Z_i - z\right), \\
\hat{f}_{X,Z}(x,z) &= \sum_{i=1}^{n} K_{X,h_X}\left(X_i - x\right)K_{Z,h_Z}\left(Z_i - z\right).
\end{aligned}$$

By standard arguments for kernel estimators (see, e.g. Li and Racine, 2006), the following holds under the smoothness assumptions imposed on the model,

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}}\left\{\frac{\partial \hat{f}_{Y,X,Z}(y|x,z)}{\partial z} - \frac{\partial f_{Y,X,Z}(y|x,z)}{\partial z} - \sum_{a \in \{Y,X,Z\}} h_a^2 B_a(y|x,z)\right\} \to^d N\left(0, \tilde{V}\left(y,x,z\right)\right), \tag{39}$$

where $B_a(y|x,z)$, $a \in \{Y, X, Z\}$, are the usual bias components due to kernel smoothing, and

$$\tilde{V}\left(y,x,z\right) = f_{Y,X,Z}(y,x,z)\int_{\mathbb{R}^{d_Y}} K_Y^2\left(y\right)dy\int_{\mathbb{R}^{d_X}} K_X^2\left(x\right)dx\int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z\left(z\right)}{\partial z}\frac{\partial K_Z\left(z\right)}{\partial z'}dz \in \mathbb{R}^{d_Z \times d_Z}. \tag{40}$$

Similarly,

$$
\begin{aligned}
\hat{f}_{Y,X,Z}(y,x,z) &= f_{Y,X,Z}(y,x,z) + O_P\left(h_x^2\right) + O_P\left(h_z^2\right) + O_P\left(\sqrt{\frac{1}{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z}}}\right), \\
\hat{f}_{X,Z}(x,z) &= f_{X,Z}(x,z) + O_P\left(h_x^2\right) + O_P\left(h_z^2\right) + O_P\left(\sqrt{\frac{1}{nh_X^{d_X}h_Z^{d_Z}}}\right), \\
\frac{\partial \hat{f}_{X,Z}(x,z)}{\partial z} &= \frac{\partial f_{X,Z}(x,z)}{\partial z} + O_P\left(h_x^2\right) + O_P\left(h_z^2\right) + O_P\left(\sqrt{\frac{1}{nh_X^{d_X}h_Z^{d_Z+2}}}\right).
\end{aligned}
$$

Under the conditions on the bandwidths, (i) all bias components are negligible and (ii) $\partial \hat{f}_{Y,X,Z}(y,x,z)/\left(\partial z\right)$ contains the leading variance terms with all other variance components being of a smaller order. This shows the first part of the lemma.

The second part follows by similar arguments with the leading term being

$$
\frac{\partial^2 \hat{f}_{Y,X,Z}(y,x,z)}{\partial z \partial z'} = \frac{\partial^2 f_{Y,X,Z}(y,x,z)}{\partial z \partial z'} + O_P\left(h_y^2\right) + O_P\left(h_x^2\right) + O_P\left(h_z^2\right) + O_P\left(\sqrt{\frac{1}{nh_y^{d_Y}h_x^{d_X}h_z^{d_Z+4}}}\right).
$$

The result now follows from the conditions on the bandwidths. The third part follows by similar arguments. ■

C Identification and estimation with discrete Z

We here consider the case where Z has a discrete, but potentially unbounded, support $\mathcal{Z}$. In this case, derivatives of $f_{\varepsilon|Z}$ w.r.t. z are not well-defined, and we therefore redefine the solution mapping Λ in terms of differences: For any collection of d_Y values $z_1, ..., z_{d_Y+1} \in \mathcal{Z}$ with $z_i \neq z_j$, $i \neq j$, let

$$
\Lambda\left(z_1, ..., z_{d_Y+1}\right) = \left\{ e \in \mathcal{E} : f_{\varepsilon|Z}\left(e|z_i\right) = f_{\varepsilon|Z}\left(e|z_j\right), \quad 1 \leq i < j \leq d_Y + 1 \right\}. \tag{41}
$$

Here, one can interpret $\left\{ f_{\varepsilon|Z}\left(e|z_i\right) - f_{\varepsilon|Z}\left(e|z_j\right) \right\} / \left\{ z_i - z_j \right\}$ as the "derivative" of the density w.r.t. z and so the above version of Λ can be thought of as a "discretized" version of the one introduced in the case of continuous Z. (See Appendix D for an illustrative example of how the solution mapping behaves) Due to the discrete nature of Z, we are only able to identify m at a discrete set of points. With a slight abuse of notation, we have

$$
\bar{m}\left(x,z\right) = m\left(x, \Lambda\left(z\right)\right), \text{ where } z \in \bar{\mathcal{E}} := \left\{ z \in \mathcal{Z}^{d_Y+1} : z_i \neq z_j \text{ for } i \neq j \right\}, \tag{42}
$$

is identified: For any such $z \in \bar{\mathcal{E}}$ and any $x \in \mathcal{X}_0$, consider a solution y^* to the following set of equations,

$$
f_{Y|X,Z}(y^*|x,z_i) = f_{Y|X,Z}(y^*|x,z_j), \quad 1 \leq i < j \leq d_Y + 1. \tag{43}
$$

From eq. (6), this set of equations is equivalent to

$$f_{\varepsilon|Z}\left(r\left(y,x\right)|z_i\right)=f_{\varepsilon|Z}\left(r\left(y,x\right)|z_j\right),\quad 1\leq i<j\leq d_Y+1, \tag{44}$$

since $\left|\partial r\left(y,x\right)/\left(\partial y\right)\right|>0$ by assumption. By the same arguments as in the case of Z being continuous, y^* satisfies $r\left(y,x\right)=e$, where $e=\Lambda\left(z\right)$, or, equivalently, $y^*\in m\left(x,\Lambda\left(z\right)\right)$. The reverse implication is easily shown to hold by analogous arguments and we conclude:

Theorem 14 *Under Assumptions 1-3: For any given $z\in\bar{\mathcal{E}}$, $\bar{m}\left(x,z\right)$, as defined in eq. (42), is identified for all $x\in\mathcal{X}_0$ as the (set of) solution(s) y^* to eq. (43).*

As in the continuous Z case, the above theorem only allows us to identify $m\left(x,e\right)$ at the values of $e\in\Lambda\left(z\right)$ for some $z\in\mathcal{Z}^{d_Y+1}$. Thus, given that $\mathcal{Z}$ is countable, we can only identify $m\left(x,e\right)$ at a countable number of values $e\in\mathcal{E}$. At the same time, for the consumers that can be reached through (43), we can identify the differences $m\left(x',e\right)-m\left(x,e\right)=\bar{m}\left(x,z\right)-\bar{m}\left(x,z\right)$, when the value of x changes to x' while the value of z stays fixed, and we can also identify marginal effects, $\partial m\left(x,e\right)/\left(\partial x\right)=\partial\bar{m}\left(x,z\right)/\left(\partial x\right)$.

Eq. (43) suggests the following nonparametric estimation strategy: Obtain a nonparametric estimator of $f_{Y|X,Z}$, say, $\hat{f}_{Y|X,Z}$, substitute this into eq. (43),

$$\hat{f}_{Y|X,Z}(y|x,z_i)=\hat{f}_{Y|X,Z}(y|x,z_j),\quad 1\leq i<j\leq d_Y+1. \tag{45}$$

and solve this w.r.t. y. As before, for the theoretical results, we here focus on the case where $\hat{f}_{Y|X,Z}$ is chosen as a kernel density estimator, which in the discrete Z case takes the form

$$\hat{f}_{Y|X,Z}(y|x,z)=\frac{\sum_{i=1}^{n}K_{Y,h_Y}\left(Y_i-y\right)K_{X,h_X}\left(X_i-x\right)\mathbb{I}\left\{Z_i=z\right\}}{\sum_{i=1}^{n}K_{X,h_X}\left(X_i-x\right)\mathbb{I}\left\{Z_i=z\right\}}. \tag{46}$$

If X has discrete components, the above estimator should be modified in the same manner as in the case of Z being continuous.

Similar to the continuous case, we can represent the estimator solving eq. (45) as GMM estimator: Let $\hat{g}(y|x,z)=\{\hat{g}_{i,j}(y|x,z):1\leq i<j\leq d_Y\}$, where $\hat{g}_{i,j}(y|x,z)=\hat{f}_{Y|X,Z}(y|x,z_i)-\hat{f}_{Y|X,Z}(y|x,z_j)$, contain the "moment restrictions" and define

$$\hat{m}\left(x,z\right)=\arg\min_{y\in\mathcal{Y}_0}\left\|\hat{g}(y|x,z)\right\|. \tag{47}$$

Note that we here do not need a weighting matrix since the moment conditions exactly identify $\bar{m}\left(x,z\right)$.

For the asymptotic analysis, we maintain Assumption 5, but can dispense of Assumption 6(ii) and 6 since these are (almost) void in the case of Z being discrete. To state the result, introduce the population version of the moment conditions, $g(y|x,z) = \{g_{i,j}(y|x,z) : 1 \leq i < j \leq d_Y\}$ where $g_{i,j}(y|x,z) = f_{Y|X,Z}(y|x,z_i) - f_{Y|X,Z}(y|x,z_j)$. We then have the following theorem whose proof is left out since it follows along the same lines as the one of Theorem 9:

Theorem 15 *Assume that Assumptions 1-4(i), 5 and 6(i) hold and* $H_m(x,z) = G(x,z)' G(x,z) \in \mathbb{R}^{d_Y \times d_Y}$ *has full rank where*

$$G(x,z) := \left. \frac{\partial g(y|x,z)}{\partial y} \right|_{y=\bar{m}(x,z)} \in \mathbb{R}^{d_Y \times d_Y}.$$

Then, for any bandwidth sequences satisfying $nh_Y^{d_Y} h_X^{d_X} h_a^4 \to 0$, *for* $a = Y, X$, *and* $nh_Y^{d_Y+2} h_X^{d_X} \to \infty$, $\hat{m}(x,z)$, *as defined by eq. (47), is consistent and satisfies*

$$\sqrt{nh_Y^{d_Y} h_X^{d_X}} \{\hat{m}(x,z) - \bar{m}(x,z)\} \to^d N\left(0, H_m^{-1}(x,z)\, G(x,z)'\, \Omega_m(x,z)\, G(x,z)\, H_m^{-1}(x,z)\right),$$

where

$$\Omega_m(x,z) = \left. \left\{ \frac{f_{Y|X,Z}(y|x,z_i)}{f_{X,Z}(x,z_i)} : 1 \leq i < j \leq d_Y + 1 \right\} \right|_{y=\bar{m}(x,z)} \int_{\mathbb{R}^{d_Y}} K_Y^2(y)\, dy \int_{\mathbb{R}^{d_X}} K_X^2(x)\, dx \in \mathbb{R}^{d_Y \times d_Y}.$$

D Multiple Identifying Sets

There may exist more than one set of variables that satisfy Assumption 3. If so, one can develop a more efficient estimator of m by combining the information contained in them. Formally, let $Z^{(k)} \in \mathbb{R}^{d_Y}$, $k = 1, ..., M$, be $M \geq 2$ distinct sets of variables satisfying:

Assumption 2* For $k = 1, ..., M$: ε is distributed independently of X conditional on $Z^{(k)}$ and $\varepsilon| \left(X, Z^{(k)}\right) = \varepsilon|Z^{(k)}$ has a continuous distribution characterized by a density $f_{\varepsilon|Z^{(k)}}\left(\varepsilon|z^{(k)}\right)$ which is twice continuously differentiable.

Assumption 3* For $k = 1, ..., M$: For any e, the following equations have a unique solution in terms of $z^{(k)}$,

$$\frac{\partial f_{\varepsilon|Z^{(k)}}\left(e|z^{(k)}\right)}{\partial z^{(k)}} = 0.$$

The solution mapping taking e into the corresponding solution $z^{(k)}$ is one-to-one.

Recall that Assumption 2 and 3 generate moment conditions which identifies $\bar{m}(x,z)$. Assumptions 2* and 3* can therefore be thought of generating over-identifying moment restrictions. Similar to Minimum Distance-estimators, these can then be combined to obtain a more efficient estimator.

We here focus on the estimation of m; the analysis of the corresponding estimator of r follows along the same lines.

Given the conditional kernel density estimators $\hat{f}_{Y|X,Z^{(k)}}(y|x,z^{(k)})$, $k=1,...,M$, we collect the M "moment conditions" in $\hat{G}(y|x,z) = (\hat{G}_1(y|x,z)', ..., \hat{G}_M(y|x,z)')' \in \mathbb{R}^{Md_Y}$ where

$$\hat{G}_k(y|x,z) := \left.\frac{\partial \hat{f}_{Y|X,Z^{(k)}}(y|x,z^{(k)})}{\partial z^{(k)}}\right|_{z^{(k)}=z}, \quad z \in \mathbb{R}^{d_Y}.$$

For a given choice of (x,z), we then propose to estimate $\bar{m}(x,z)$ by

$$\hat{m}(x,z) = \arg\min_{y\in\mathbb{R}^{d_Y}} \hat{G}(y|x,z)' \hat{W}(x,z) \hat{G}(y|x,z), \tag{48}$$

for some weighting matrix $\hat{W}(x,e) \in \mathbb{R}^{Md_Y\times Md_Y}$. To state the limiting distribution of the estimator, we define $G(y|x,z) = (G_1(y|x,z)', ..., G_M(y|x,z)')' \in \mathbb{R}^{Md_Y}$ where

$$G_k(y|x,z) := \left.\frac{\partial f_{Y|X,Z^{(k)}}(y|x,z^{(k)})}{\partial z^{(k)}}\right|_{z^{(k)}=z}.$$

The following theorem generalizes Theorem 9, where, for simplicity, we assume that the same bandwidths is used across the M density estimates:

Theorem 16 *Suppose that Assumptions 1, 2*-3* and 4-5 hold, and the matrix*

$$H(x,z) := G_y(x,z)' G_y(x,z) \in \mathbb{R}^{d_Y\times d_Y}$$

has full rank, where $G_y(x,e) := \partial G(y|x,\varepsilon)/(\partial y')|_{y=\bar{m}(x,z)} \in \mathbb{R}^{Md_Y\times d_Y}$*. Then, for any bandwidth sequences satisfying* $nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Y+2}h_a^4 \to 0$ *for* $a=Y,X,Z$, $nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Y+4}\to\infty$ *and* $nh_Y^{d_Y+2}h_X^{d_w}h_Z^{d_Y+2}\to\infty$, $\hat{m}(x,z)$*, as defined by eq. (48), is consistent and satisfies*

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Y+2}}\{\hat{m}(x,z)-\bar{m}(x,z)\} \to^d N(0,\Omega(w,z)),$$

where

$$\Omega(x,z) = H^{-1}(x,z) G_y(x,z)' W(x,z) V(x,z) W(x,z) G_y(x,z) H^{-1}(x,z),$$

and $V(x,z) = [V_{ij}(x,z)]_{i,j=1}^M$ *with*

$$V_{ii}(x,z) = \left.\frac{f_{Y|X,Z^{(i)}}(y^*|x,z)}{f_{X,Z^{(i)}}(x,z)}\right|_{y^*=\bar{m}(x,z)} \int_{\mathbb{R}^{d_Y}} K_Y^2(y)\,dy \int_{\mathbb{R}^{d_X}} K_X^2(x)\,dx \int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z(z)}{\partial z}\frac{\partial K_Z(z)}{\partial z'}dz \in \mathbb{R}^{d_Y\times d_Y},$$

and, for $i\neq j$,

$$V_{ij}(x,z) = \left.\frac{f_{Y|X,Z^{(i)},Z^{(j)}}(y^*|x,z,z)}{f_{X,Z^{(i)},Z^{(j)}}(x,z,z)}\right|_{y^*=\bar{m}(x,z)} \int_{\mathbb{R}^{d_Y}} K_Y^2(y)\,dy \int_{\mathbb{R}^{d_X}} K_X^2(x)\,dx \int_{\mathbb{R}^{d_Z}} \frac{\partial K_Z(z)}{\partial z}\frac{\partial K_Z(z)}{\partial z'}dz \in \mathbb{R}^{d_Y\times d_Y}.$$

In the case where $M \geq 2$, we can use a J-test to test for whether the chosen Z's indeed are valid co-variates satisfying Assumptions 2*–3*. As is standard for Minimum Distance-type estimators, an efficient estimator arises by choosing $\hat{W}(x,z)$ to be a consistent estimator of $W(x,z) = V^{-1}(x,z)$ in which case the asymptotic variance of $\hat{m}(x,e)$ takes the form $\Omega(w,z) = \left[Gy(x,z)' V^{-1}(x,z) G_y(x,z)\right]^{-1}$.

E Simulation Study

Here we investigate the performance of the estimator for the vector of characteristics Z^* corresponding to a value of Y for different values of X. The data-generating process is chosen as a bivariate ($d_Y = 2$) random coefficient model where

$$Y_k = X\varepsilon_k, \quad \text{and } \varepsilon_k = Z_k + \eta_k,$$

for $k = 1, 2$. In total,

$$Y = ZX + X\eta.$$

We assume X, Z, and η are mutually independent with $\eta \sim N(\mu_\eta, \Omega_\eta)$. Thus, $\varepsilon|Z \sim N(Z + \mu_\eta, \Omega_\eta)$ and $Y|(X,Z) \sim N\left(ZX + \mu_\eta, x^2\Omega_\eta\right)$. As such its density is given by

$$f_{Y|X,Z}(y|x,z) = \frac{1}{\sqrt{(2\pi)^d \Sigma(x)}} \exp\left\{-\frac{1}{2}(y - xz)' \Sigma^{-1}(x)(y - xz)\right\},$$

where $\Sigma(x) = x^2\Omega_\eta$. In particular,

$$\hat{z}(y,x) := \arg\max_z f_{Y|X,Z}(y|x,z) = \frac{y}{x},$$

which is the inverse $r(y,x) = y/x$ of the structural relation $Y = m(X,\varepsilon) = X\varepsilon$. For given values of (y,x), we implement the estimator of $r(y,x)$ defined as $\hat{r}(y,x) = \arg\max_z \hat{f}_{Y|X,Z}(y|x,z)$ where $\hat{f}_{Y|X,Z}(y|x,z)$ is the kernel estimator of the conditional density using a matrix of bandwidths, H. The bandwidth matrix are chosen using the multivariate version of Silverman's Rule-of-Thumb,

$$H = n^{-1/(2d_Y+1)}\hat{\Sigma}^{1/2},$$

where $\hat{\Sigma}$ is the sample covariance matrix of (Y, X, Z).

The results for the estimator $\hat{r}(y,x) = (\hat{r}_1(y,x), \hat{r}_2(y,x))$ are reported in the following Figures. In each figure we fix y at a particular value, say, $\overline{y}$, and then plot the estimates of the function $x \mapsto r_1(x,\overline{y})$ and $x \mapsto r_2(x,\overline{y})$. The results show that the kernel-based estimator works quite well, with small biases and not too big variances.

Figure 3: Estimation of $r_1(x, y)$ with y fixed.

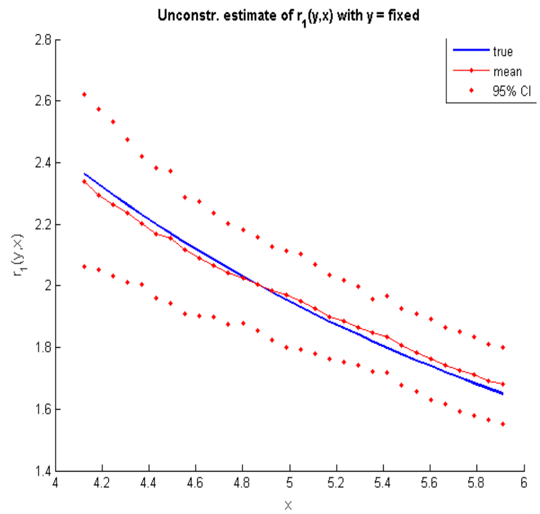
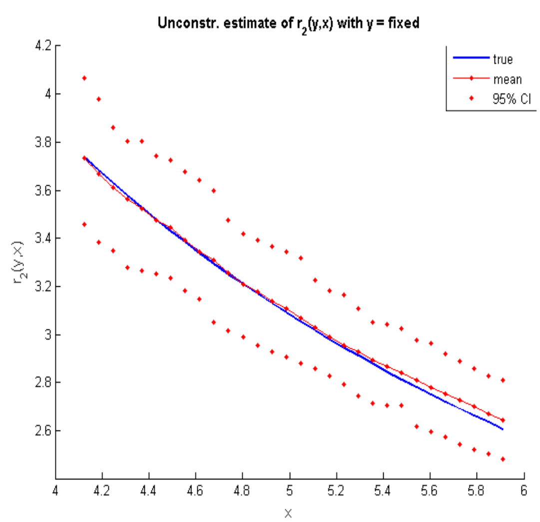


Figure 4: Estimation of $r_2(x, y)$ with y fixed.



F Computation

To calcualte the estimators, we used a grid search algorithm. Given x^* characterizing a budget and a $z^* = (z_1^*, z_2^*)$, the values of $y = (y_1, y_2)$ minimizing

$$\left(\left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=z^*}(y_1, y_2)}{\partial z_1}\right|\right)^2 + \left(\left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=z^*}(y_1, y_2)}{\partial z_2}\right|\right)^2$$

over the set where $\left|\widehat{f}_{Y|X=x^*,Z=z^*}(y_1, y_2)\right|$ and $\left|\widehat{f}_{Y,X,Z}(y_1, y_2, x^*, z^*)\right|$ are sufficiently away from zero were found by grid and subgrid searches. Specifically, let a,b be constants. Let $y_1^{(1)}, y_1^{(2)}, ..., y_1^{(L_1)}$ denote increasing values of y_1 forming a grid over the support of Y_1. Let $y_2^{(1)}, y_2^{(2)}, ..., y_2^{(L_2)}$ denote increasing values of y_2 forming a grid over the support of Y_2. If for k_1 and k_1+1 and for k_2 and k_2+1, it is the case that

$$\max\left\{\left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1)}, y_2^{(k_2)}\right)}{\partial z_j}\right|_{j=1,2}, \left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1+1)}, y_2^{(k_2)}\right)}{\partial z_j}\right|_{j=1,2}\right\} \leq a$$

and

$$\max\left\{\left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1)}, y_2^{(k_2+1)}\right)}{\partial z_j}\right|_{j=1,2}, \left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1+1)}, y_2^{(k_2+1)}\right)}{\partial z_j}\right|_{j=1,2}\right\} \leq a$$

while

$$\min\left\{\left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1)}, y_2^{(k_2)}\right)\right|, \left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1+1)}, y_2^{(k_2)}\right)\right|\right\} \geq b$$

$$\min\left\{\left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1)}, y_2^{(k_2+1)}\right)\right|, \left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1+1)}, y_2^{(k_2+1)}\right)\right|\right\} \geq b$$

$$\min\left\{\left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1)}, y_2^{(k_2)}, x^*, z^*\right)\right|, \left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1+1)}, y_2^{(k_2)}, x^*, z^*\right)\right|\right\} \geq b$$

and

$$\min\left\{\left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1)}, y_2^{(k_2+1)}, x^*, z^*\right)\right|, \left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1+1)}, y_2^{(k_2+1)}\right)\right|\right\} \geq b$$

then, the rectangle is itself subdivided, forming a grid, with $y_1^{(k1)} = y_1^{(k1,1)}, y_1^{(k_1,2)}, ..., y_1^{(k_1,L_1)} = y_1^{(k_1+1)}$ and $y_2^{(k2)} = y_2^{(k2,1)}, y_2^{(k_2,2)}, ..., y_2^{(k_2,L_2)} = y_2^{(k_2+1)}$. The program then checks for all points within the subgrid, indexed as (k_1, j_1), (k_1, j_1+1) and (k_2, j_2), (k_2, j_2+1) for whether

$$\max\left\{\left|\frac{\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1)},y_2^{(k_2,j_2)}\right)}{\partial z_j}\right|_{j=1,2},\left|\frac{\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1+1)},y_2^{(k_2,j_2)}\right)}{\partial z_j}\right|_{j=1,2}\right\}\leq a\ /2$$

and

$$\max\left\{\left|\frac{\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y(_1^{(k_1,j_1)},y_2^{(k_2,j_2+1)}\right)}{\partial z_j}\right|_{j=1,2},\left|\frac{\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1+1)},y_2^{(k_2,j_2+1)}\right)}{\partial z_j}\right|_{j=1,2}\right\}\leq a/2$$

while

$$\min\left\{\left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1)},y_2^{(k_2,j_2)}\right)\right|_{j=1,2},\ \left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1+1)},y_2^{(k_2,j_2)}\right)\right|_{j=1,2}\right\}\geq 2\ b$$

$$\min\left\{\left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1)},y_2^{(k_2,j_2+1)}\right)\right|_{j=1,2},\ \left|\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(k_1,j_1+1)},y_2^{(k_2,j_2+1)}\right)\right|_{j=1,2}\right\}\geq 2\ b$$

$$\min\left\{\left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1,j_1)},y_2^{(k_2,j_2)},x^*,z^*\right)\right|,\ \left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1,j_1+1)},y_2^{(k_2,j_2)},x^*,z^*\right)\right|\right\}\geq 2\ b$$

and

$$\min\left\{\left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1,j_1)},y_2^{(k_2,j_2+1)},x^*,z^*\right)\right|,\ \left|\widehat{f}_{Y,X,Z}\left(y_1^{(k_1,j_1+1)},y_2^{(k_2,j_2+1)}\right)\right|\right\}\geq 2\ b$$

This procedure can be continued by subdividing again at the same time as lowering the upper bound for $\left|\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y_1^{(\cdot)},y_2^{(\cdot)}\right)/\partial z_j\right|$. The point with the lowest values for

$$\left(\left|\frac{\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y(_1^{(k_1,j_1)},y_2^{(k_2,j_2+1)}\right)}{\partial z_1}\right|\right)^2+\left(\left|\frac{\partial\widehat{f}_{Y|X=x^*,Z=z^*}\left(y(_1^{(k_1,j_1)},y_2^{(k_2,j_2+1)}\right)}{\partial z_2}\right|\right)^2$$

among those at the end points of grids is the one reported as the solution.

A similar procedure was used when given a budget characterized by x^* and a point $y^*=(y_1^*,y_2^*)$, the search is for (z_1^*,z_2^*) that minimizes

$$\left(\left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=(z_1,z_2)}\left(y_1^*,y_2^*\right)}{\partial z_1}\right|\right)^2+\left(\left|\frac{\partial \widehat{f}_{Y|X=x^*,Z=(z_1,z_2)}\left(y_1^*,y_2^*\right)}{\partial z_2}\right|\right)^2$$

over the set of $z=(z_1,z_2)$ where $\left|\widehat{f}_{Y|X=x^*,Z=z}\left(y_1^*,y_2^*\right)\right|$, $\left|\widehat{f}_{Y,X,Z}\left(y_1^*,y_2^*,x^*,z\right)\right|$, and $\left|\widehat{f}_{X,Z}\left(x^*,z\right)\right|$ are sufficiently away from zero.