cemmap

centre for microdata methods and practice

Individual counterfactuals with multidimensional unobserved heterogeneity

Richard Blundell
Dennis Kristensen
Rosa Matzkin

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP60/17

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

An ESRC Research Centre

Individual Counterfactuals with Multidimensional Unobserved Heterogeneity*

Richard Blundell† Dennis Kristensen‡ Rosa Matzkin§

First Draft: June 2014. This Draft: December 2017

Abstract

New nonparametric methods that identify and estimate counterfactuals for individuals, when each is characterized by a vector of unobserved characteristics, are developed and applied to estimate systems of individual consumer demand and welfare measures. The unobserved characteristics are allowed to enter in unrestricted ways. Identification is delivered through two fundamental assumptions: First, the system is invertible in the vector of unobserved heterogeneity. Second, there exist external, individual-specific, covariates that are related to the unobserved heterogeneity and do not enter directly into the system of interest. The observed external variables can be either discrete or continuously distributed. Estimators based on the identifying restrictions are developed and their asymptotic properties derived. Using UK micro data on consumer demand, we apply the methods to estimate individual demand counterfactuals subject to revealed preference inequalities.

JEL: C20, D12

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

---

*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, Jim Heckman, Stefan Hoderlein, Arthur Lewbel, and Whitney Newey, among others, as well as seminar participants at many seminars and conferences, including Harvard/MIT, 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, 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 R000239865) is gratefully acknowleged. Blundell acknowledges support from 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 FES made available by the ONS through the ESRC Data Archive has been used by permission of the controller of HMSO. Neither the ONS nor the ESRC Data Archive bear responsibility for the analysis or the interpretation of the data reported here. 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

Standard theoretical models of individual behavior, such as those of a consumer or a firm, are usually developed under the assumption that the behavior of the individual is generated by an optimization problem whose objective function is fixed across different values of exogenous variables. Analyzing such models with data requires either to observe each individual repeatedly or, if the data is from repeated cross-sections, to assume that all observed individuals have some common objective function component. In this paper, we develop a new method for analyzing individual behavior using repeated cross-sections but without assuming common elements at the level of the objective functions of the individuals. We let the objective function of each individual be random, in the sense that it depends in an unrestricted manner, with no functional form restrictions, on a vector of unobservable characteristics. Under the assumption that the distribution of each cross-section is generated from the distribution of this vector of unobservable variables, our method allows one to analyze the behavior of each individual, as if the objective function of each individual were known and fixed. Our results can be applied to estimate counterfactuals for individual behavior and welfare comparisons, when each individual is characterized by a vector of unobserved characteristics, in a number of important settings including equilibrium models of differentiated products, models of hedonic equilibrium, and models of multidimensional optimization.

The general model we consider is a nonparametric system of reduced-form equations, where each endogenous variable is an unknown function of a set of observed exogenous covariates and a vector of unobserved variables, the latter characterizing the "multidimensional unobserved heterogeneity". In these, our identification strategy for the functions defining the structural relationship relies on two fundamental assumptions: First, the system of functions is invertible in the vector of unobserved heterogeneity. Second, we have observed additional individual-specific covariates that are related to the unobserved heterogeneity, but do not enter directly into the set of simultaneous equations generating the system of reduced form equations. We henceforth refer to these as "external" covariates. They can be interpreted as observable proxies or as measurements of the latent variables. The external covariates can be strong or weak representatives of the unobservables, depending on whether they characterize the unobservables uniquely or not. When the external covariates are continuously distributed and the dependence between and the unobservables is strong enough, one can trace the behavior of the endogenous variables across different values of the internal covariates keeping the value of the unobservables fixed. In such case, each function in the system is globally nonparametrically identified. If the representation is weak, in the sense that the distribution of the external variables do not uniquely characterize the one of the unobservables, we can only trace the behavior partially, providing sets of values for the endogenous variables. When the external covariates are discrete, we are only able to identify the functions defining the system of equations at particular values of unobserved heterogeneity, and depending on the strength of the representation, we can trace out either a function or a correspondence, both across different values of the internal covariates. Whether the external covariates represent the unobservables uniquely or not is testable.

Our identification results are constructive in the sense that they lead to natural estimators of the unknown functions of interest. We develop nonparametric estimators of the functions and analyze their asymptotic properties. The estimators and their asymptotic theory are in turn used to carry out inference on individual responses to counterfactual changes in the observed internal covariates, when the value of the unobservable variables is fixed. We also show that the performance of the estimators can be improved in a number of relevant scenarios.

Other identification and/or estimation methods for nonparametric models with multivariate unobserved heterogeneity exist. However, these either impose additional functional form restrictions, which may be violated in data, or are not able to identify counterfactual responses. Papers dealing within the first category include, amongst others, Matzkin (2003, 2008, 2015), Berry and Haile (2014, 2015), Carlier, Chernozhukov and Galichon (2016), and Lewbel and Pendakur (2017). These effectively restrict how unobserved heterogeneity enters the system of equations of interest. For example, Matzkin (2003, Appendix A) and Lewbel and Pendakur (2017) assume knowledge of a function that depends on observable and unobservable variables. Carlier, Chernozhukov and Galichon (2016) require that the functions of interest are given as partial derivatives of a convex function, thereby implicitly imposing a symmetry restriction on the functional form. Example 4.2 in Matzkin (2008), the estimators in Matzkin (2015), and the models in Berry and Haile (2014, 2015) assume that at least one of the unobserved heterogeneity terms enters the system in form of an index, thereby again imposing functional form restrictions. In terms of the second category, it is well-known that one can always represent the distribution of outcome variables in terms of a nonparametric function and a nonparametric distribution of unobservable variables in such a way that for each value of the vector of endogenous variables one can point at the value of the vector of unobservable variables corresponding to it. However, in general these representations cannot be used for individual counterfactuals where the vector of unobserved heterogeneity has to remain fixed while the conditioning variables vary. (See Benkard and Berry, 2006; Matzkin, 2008). Thus, these results do not allow for identification of counterfactual responses.

Compared to the existing literature, our main contribution is to show that functional form restrictions can be replaced by a weak additional model for the unobservables: Specifically, to achieve identification we require external covariates that only effect the outcome variables through the unobserved heterogeneity. This idea is not new to the literature on structural modelling and is often used in empirical work, where individual heterogeneity (random coefficients) are explicitly modelled as functions of observed characteristics and random shocks; see, e.g., Berry, Levinsohn and Pakes (2004). Our identfication strategy also shares features with Cunha, Heckman and Schennach (2010) where measurements are also used to identify latent factors. However, to our knowledge, this approach has not been employed in a nonparametric setting before, and we are able to demonstrate its usefulness in flexible identification and estimation of simultaneous equations models.

Our methods can be applied to consumer demand data with continuous or discrete budget variation. In our empirical application, we consider a nonparametric model for the demand for a vector of goods by individual consumers characterized by a vector of prices, income and unobserved tastes. The external covariates needed for identification are chosen from an observed set of household characteristics that seem reasonable to assume being correlated with the consumers' unobserved preferences. For each of a finite number of observed vectors of prices, and for continuous income levels, we then show identification of the demand function for any given consumer in the population, and conduct counterfactual inference for consumer responses to income and price changes. Given that we only have available a finite number of price regimes in our data set, we are unable to point identify counterfactual demand for prices that have not been seen before. Instead we impose revealed preference inequality restrictions on the demands generated by each vector of unobserved tastes, which allows us to compute sharp and informative bounds on responses to counterfactual price changes.

Our empirical application can be thought of as a generalization of Blundell, Browning, and Crawford (2003, 2008), who worked within the same framework as we do to identify consumer demand and welfare counterfactuals. However, in their work it is assumed that, on each budget set, the distribution of demand was generated from the demand of a single consumer with an additive (measurement) error and with revealed preference restrictions on the demand of the single consumer.[1] In contrast, we assume that the distribution of demand is generated from a distribution of unobserved tastes. Another related paper is Blundell, Kristensen, and Matzkin (2014) who studied revealed preference restrictions using conditional quantiles in a model where the distribution of demand is generated by a distribution of tastes. Their method was restricted to the case of a single good where demand and taste were scalars, and demand was increasing in taste. Blundell, Horowitz and

[1] The generalization of Blundell, Browning and Crawford (2008) to derive sharp bounds for the strong axiom SARP inequalities, rather than WARP, is given in Blundell, Browning, Cherchye, Crawford, de Rock, and Vermeulen (2015).

Parey (2016) consider the case of scalar demand and scalar heterogeneity with continuous prices. Hausman and Newey (2016, 2017) consider a single demand equation with continuous prices and multiple unobserved taste heterogeneity; using bounds on income responses, they show that without further assumptions identification is limited to bounds on average welfare measures. Estimation of individual, rather than average, welfare measures from demands with multidimensional taste vectors requires methods appropriate for estimation of systems of simultaneous equations, as developed in this paper.

Invertibility of consumer demand, which is one of the key assumptions for our identification strategy, is guaranteed by imposing conditions on the random utility function of consumers, generalizing the results in Brown and Matzkin (1998) and Beckert and Blundell (2008). Berry, Gandhi, and Haile (2013) and Chiappori, Komunjer and Kristensen (2016a) provide invertibility conditions on demand functions that could also be used. Employing a stochastic revealed preference approach, as in McFadden (2005), Kitamura and Stoye (2014) avoid having to impose invertibility but are on the other hand only able to identify distributions of demands. In contrast, our approach provides counterfactuals for any given individual in the distribution. For scalar unobserved heterogeneity, invertibility follows when the demand function is increasing in heterogeneity. In the case of two goods, Hoderlein and Stoye (2015) show that rationalization with an invertibility assumption is without loss of generality. Hoderlein and Stoye (2014) use the same assumptions and show how to bound from above and below the fraction of the population who violate the weak axiom WARP.

We apply our framework to the prediction of demand by consumers in the UK. We use the history of the Family Expenditure Survey which provides consumer expenditure data at the household level for a large representative sample of consumers in the UK. We estimate counterfactual demands for food, services and other goods for a consumer characterized by a given choice of total budget and unobserved heterogeneity. The estimates suggest a mildly downward sloping own demand for food and a upward sloping cross-demand curve for food with respect to the price of services. The results show that the bounds on counterfactual demands on demands at previously unobserved prices can be quite tight, especially where the data is dense, becoming wider where there is sparse data.

The remainder of the paper is as follows. We present the multidimensional framework in the next section. Conditions for identification are presented in Sections 3. In section 4 we develop nonparametric estimators and discuss various extensions in section 5. In section 6 we consider how the proposed identification and estimation strategy can be used in for individual consumer demand and welfare counterfactuals. Section 7 examines implementation and finite sample performance. The estimated bounds on counterfactual demands for the British consumer micro-data are presented in section 8. Section 9 concludes.

2 Nonparametric Models with Multidimensional Heterogeneity

We consider a model specified as

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

where the random vectors $Y = (Y_1, ..., Y_{d_Y})'$ and $X = (X_1, ..., X_{d_X})'$ are observed, while the random vector $\varepsilon = (\varepsilon_1, ..., \varepsilon_{d_\varepsilon})'$ is unobserved. Both ε and Y are assumed to be continuously distributed. In the following, let $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$, $\mathcal{X} \subseteq \mathbb{R}^{d_X}$ and $\mathcal{E} \subseteq \mathbb{R}^{d_\varepsilon}$ denote the supports of Y, X and ε, respectively. Given data of Y and X, we then wish to identify the vector-valued function $m : \mathcal{X} \times \mathcal{E} \mapsto \mathcal{Y}$ and the distribution of ε.

Existing results on nonparametric identification of models on the form (1) impose strong restrictions on the functional form of m. As an example, some results impose the restriction that $m(x, e) = (m_1(x, e_1), ..., m_{d_Y}(x, e_{d_Y}))'$ so that the gth component, m_g, only depends on a scalar unobservable, ε_g, $g = 1, ..., d_Y$. If, furthermore, ε is independent of X, and $e_g \mapsto m_g(x, e_g)$ is monotone, then identification of $m_g(x, e_g)$ can be achieved up to a monotone transformation of ε_g, , $g = 1, ..., d_Y$, c.f. Matzkin (2003). However, in the general case, where each component m_g is a function of multiple components of ε, additional restrictions are required in order to identify m. 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. Matzkin (2008) developed necessary and sufficient conditions for observational equivalence and provided examples of identified models subject to functional form restrictions. Further examples of identification under functional form restrictions include Matzkin (2015), Berry and Haile (2014, 2015), Carlier, Chernozhukov and Galichon (2016), and Chiappori, Komunjer and Kristensen (2016a,b).

In this paper, we develop nonparametric identification and estimation results of m under weak requirements on its functional form; see below for the precise conditions. Specifically, we replace functional form restrictions, as used in the above cited papers, by additional requirements on data availability: The key ingredient in our identification argument is the availability of additional d_Z observed variables, that we collect in 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), and so Z only affects Y through ε. The covariation between Z and ε will then allow us to use the observed variation in Z to learn about the function m and the distribution of ε.[2]

Our identification results and estimation methods for features of the function $m(x,e)$ will be developed at particular values of x and e. Hence, some of our assumptions are only required to be satisfied on a subset of the support of X.

Assumption 1. For any given x in some subset $\mathcal{X}_0 \subseteq \mathcal{X}$, the function $e \mapsto m(x,e)$ is continuously differentiable with inverse $r(x,y)$. That is, for any $x \in \mathcal{X}_0$,

$$y = m(x,e) \Leftrightarrow e = r(x,y).$$

Assumption 2. ε is distributed independently of X conditional on Z so that $\varepsilon|(X,Z) \overset{d}{=} \varepsilon|Z$ has a continuous distribution characterized by a density $f_{\varepsilon|Z}(e|z)$ which is twice continuously differentiable.

Assumption 1, or variations of it, are commonly met in the literature on nonparametric identification of simultaneous equations; see, e.g., Matzkin (2008, 2015), Berry and Haile (2014, 2015), Chiappori, Komunjer and Kristensen (2016a,b). Note that Assumption 1 implicitly restricts the unobserved variables, ε, to be of the same dimension as Y, $d_Y = d_\varepsilon$. However, our modelling framework allows for ε to be a function of a potentially high-dimensional unobservable vector, η, together with the external covariates Z; this is similar to the ideas found in Appendix A in Matzkin (2003), Matzkin (2012), and Lewbel and Pendakur (2017).

Assumption 2 restricts the distribution of unobservables to be continuous and the internal covariates, X, to be exogenous conditional on the external ones, Z. Again, assumptions of this type are typical in the literature on nonparametric identification. One can think of Z being a set of control variables, as in Imbens and Newey (2009) and Blundell and Matzkin (2014). In this sense, certain types of endogeneity of X is allowed for. However, Z will play an additional key role as a type of either a stochastic proxy for a stochastic measurement for ε as explained below, which is the novel part of our modelling approach. Also note that while we restrict ε to be continuous, we do not require X to be continuous. Similarly, at this stage we have made no assumptions about the precise nature of Z. It may have both discrete and continuous components, and it may be correlated with X. Later, however, as already noted, we will impose further restrictions on the stochastic relationship between Z and ε to achieve identification.

[2] A related idea can be found in Chiappori, Komunjer and Kristensen (2016b) where identification is also achieved by imposing restrictions on the distribution of unobservables; but their restrictions are of a very different nature to the one developed here.

Under Assumption 1, an equivalent representation of eq. (1) is

$$r(Y,X) = \varepsilon,$$

for $X \in \mathcal{X}_0$. Assumptions 1-2 then imply that $Y|(X,Z) = (x,z)$, with $x \in \mathcal{X}_0$, has a continuous distribution as represented by the following conditional density,

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

where $|\partial r(x,y)/\partial y|$ denotes the absolute value of the Jacobian determinant of $r(x,y)$ with respect to y. The left hand side in this equation can be estimated. The right hand side involves the function r and the conditional density of ε given Z, both of which are unknown.

2.1 Example: Consumer Demand

To focus on one modeling framework when developing the methodology, we consider a consumer characterized by income level $I \in \mathbb{R}_+$ (representing the total budget available to the consumer potentially adjusted for durable good expenditures and saving) together with observed and unobserved individual characteristics which we collect in $W \in \mathcal{W}$ and $\varepsilon \in \mathcal{E}$, respectively. 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,

$$Y = m(p, I, W, \varepsilon). \tag{3}$$

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

$$U(Y,\varepsilon) = \sum_{g=1}^{d_Y} \varepsilon_g \ln(Y_g) + Y_{d_Y+1},$$

where Y_{d_Y+1} is the “residual good". The corresponding system of demand functions is such that 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 parametric specification of preferences being correct. If not, 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 next section, this can be achieved at an individual level. Consider a consumer characterized by $(\varepsilon, W) = (e, w)$ who receives an income and price shock changing income $I = i_0$ and prices $p = p_0$ to i_1 and p_1, respectively. The consumer's demand response is $\Delta y = m(p_1, i_1, w, e) - m(p_0, i_0, w, e)$. Our identification argument will allow us to identify Δy and well as marginal effects such as $\partial m(p, I, w, e) / (\partial I)$ when I is continuously distributed and $\partial m(p, I, w, e) / (\partial p)$ when p is continuously distributed. The value of ε at which the counterfactuals will be identified can be defined from the initial observed demand $y_0 = m(p_0, i_0, w, e)$. 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.

The above demand analysis is feasible within our framework under weak additional regularity conditions on the demand model. Consider first Assumption 1: As mentioned above, 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 next proposition provides a practical method for incorporating multidimensional unobserved heterogeneity around commonly used deterministic utility functions, in a way that generates invertible demand function. The result generalizes the conditions on utility functions shown in Brown and Matzkin (1998) and Beckert and Blundell (2008) to generate invertible demands. We here 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.

Proposition 1 *Suppose that the utility function* $U(y, y_{d_Y+1}, \varepsilon) = U_1(y, y_{d_Y+1}) + U_2(y, \varepsilon)$ *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*.*

Proof. Since U_1 and U_2 are strictly monotone in (y, y_{d_Y+1}) 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 $s(y,p,I,\varepsilon)$ denote the vector of the d_Y functions such that $s(y,p,I,\varepsilon) = 0$ denotes this system of first order conditions, and let $y = m(p,I,\varepsilon)$ denote the demand function which satisfies $s(m(p,I,\varepsilon), p, I, \varepsilon) = 0$. We will show that for each (y,p,I), $\varepsilon \mapsto s(y,p,I,\varepsilon)$ is globally univalent (see Gale and Nikaido, 1965). This guarantees the global existence of an implicit function $r(y,p,I)$ such that for all (y,p,I) in a region, $s(y,p,I,r(y,p,I)) = 0$. The uniqueness of m on ε and of r on y, for any (p,I), imply that

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

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. ■

Assumption 2 requires that, in addition to W, we have observed a set of consumer-specific covariates Z which covary with ε and do not enter directly (conditional on ε) into the demand function (3). Suppose that we observe a number of individual characteristics for each consumer. We will then split the set of characteristics into two groups: The first group of characteristics is included in 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. These observed characteristics are in this sense absorbed into the unobserved component ε, and so we do not control for the effect of the second set of characteristics on demand explicitly. At this stage, no requirements on the relationship between the second set of characteristics and the unobserved components are imposed, and so we are free to assign a given characteristics to either W or Z. However, we will later on require that Z covary with ε and so the particular characteristics assigned to Z will need to be related to the taste preferences characterizing the consumer. As Z shifts around, ε has also to move. In our empirical application, Z is computed as an index of household members' age, birth cohort and education profile; it seems plausible that these variables affect consumer preferences.

Third, we require that $X \perp \varepsilon | Z$ with X defined earlier. One will in general expect income I to be endogenous, and that certain observed characteristics comove with the unobserved components. Here, Z may then play a double role: First, as control variable which ensures that I is exogenous conditional on Z; second, as a proxy for the unobserved characteristics.

3 (Partial) Identification of Counterfactuals

In this section, we impose restrictions on the stochastic relationship between Z and ε, which allows us to identify counterfactual such as the effect of a discrete change in the internal variables, $m(x,\varepsilon) - m(x',\varepsilon)$, the effect of an infinitesimal change in x, $\partial m(x,\varepsilon)/\partial x$, or the function $m(x,\varepsilon)$. The main idea is the following: Since Z does enter the structural model (1) explicitly, we know that Z only affects Y through ε. Formally, from eq. (2), we know that variation in $f_{Y|X,Z}(y|x,z)$, as we move z around, must come through changes in the second argument of $f_{\varepsilon|Z}(r(x,y)|z)$. Depending on the strength of the dependency between ε and Z, as measured by $f_{\varepsilon|Z}(e|z)$, we can then either partially or point identify the mapping $m(x,e)$.

Our identification argument proceeds in two steps: We first provide a general partial identification result that holds without further restrictions on the model besides Assumptions 1-2 above. To achieve point identification, we then impose further restrictions on the relationship between ε and Z. Specifically, if Z is continuous and $f_{\varepsilon|Z}(e|z)$ satisfies a uniqueness condition, which is testable (see below for the precise condition), we can point identify for each z in the support of Z, the value of $m(x,\varepsilon)$ at a particular ε characterized by z. If, furthermore, $f_{\varepsilon|Z}(e|z)$ satisfies an invertibility condition, which is also testable, we can trace the function $m(x,\varepsilon)$ for ε characterized as the unique value satisfying $y = m(x,\varepsilon)$. If, on the other hand, Z is discrete then point identification can only be achieved for some values of ε. In the following two subsections, we treat the cases of Z being continuously and discretely distributed, respectively.

3.1 Identification with Continuous Z

In the case where Z is continuous, we use the "score function" $\partial \log f_{\varepsilon|Z}(e|z)/(\partial z)$ to measure the dependence between ε and Z: For any given value of $z \in \mathcal{Z}$, we consider solutions e to the following set of score equations:

$$\frac{\partial \log f_{\varepsilon|Z}(e|z)}{\partial z} = 0 \text{ (or equivalently } \frac{\partial f_{\varepsilon|Z}(e|z)}{\partial z} = 0). \tag{4}$$

Depending on the stochastic relationship between ε and Z, there may be a unique solution to the above set of equations, or multiple ones. To allow for multiple solutions, we introduce the associated solution mapping,

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

where we set $\bar{\Lambda}(z) = \varnothing$ if no solution exists. In general, Λ is a set-valued mapping that measures the representation strength between Z and ε with the volume/radius of the set $\Lambda(z)$ being inversely related to the level of representation between Z and ε: At one extreme, suppose that this relationship, as measured by the conditional density $f_{\varepsilon|Z}$, is strong enough so that there is a unique solution to the score equations, then $\Lambda(z)$ is a singleton and its volume is zero. In order for $\Lambda(z)$ to be a singleton, 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. Then $f_{\varepsilon|Z}(e|z) = f_\varepsilon(e)$ and $\Lambda(z) = \mathcal{E}$ which is the maximum volume that it can achieve. Finally, 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 charaterized 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 the individual's particular value of ε. But Λ does provide information about the distributional relationship, and this will suffice for our identification results.

The dependence structure between ε and Z, as measured by Λ, then allows us to (partially) identify r (and therfore m) through the following set of "moment" conditions,

$$\frac{\partial f_{Y|X,Z}(y|x,z)}{\partial z} = \frac{f_{\varepsilon|Z}(r(x,y)|z)}{\partial z} \left| \frac{\partial r(x,y)}{\partial y} \right| = 0. \tag{6}$$

The set of solutions to (6) w.r.t. y will rely on the properties of Λ. If $\Lambda(z)$ is a singleton, we will be able to identify $m(x,e)$, and therefore also counterfactuals such as $m(x',e) - m(x,e)$. If $\Lambda(z)$ is a set, we will only be able to identify a set that $m(x,e)$ belongs, and so only partially identify counterfactuals on the form $m(x',e) - m(x,e)$. Either result will hold without further restrictions on Λ or specification of the distribution of ε.

However, without knowledge of Λ, we will not be able to identify the distribution of ε and so can only identify m up to a normalization. To see this and also to motivate notation used in the following, suppose for the moment that Λ is in fact one–to–one. Then an observational equivalent representation of the model is $Y = \bar{m}(X, \bar{\varepsilon})$ where $\bar{m}(x, \bar{e}) = m(x, \Lambda(\bar{e}))$ and $\bar{\varepsilon} = \Lambda^{-1}(\varepsilon)$. In this case we can identify the function $\bar{e} \mapsto \bar{m}(x, \bar{e})$ and the distribution of $\bar{\varepsilon}$.

Returning to the general case where Λ is not one-to-one and set-valued, we will only be able to identify the set-valued function

$$\bar{m}(x, \bar{e}) := m(x, \Lambda(\bar{e})) = \{ y \in \mathcal{Y} : y = m(x,e) \text{ for some } e \in \Lambda(\bar{e}) \}, \tag{7}$$

where we here and in the following use $\bar{e}$ (instead of z) as argument of Λ and $\bar{m}$ to avoid confusing $\bar{\varepsilon}$ and Z with each other. The set $\bar{m}(x, \bar{e})$ contains the outcomes within the group of individuals characterized by $X = x$ and $\varepsilon \in \Lambda(\bar{e})$. We show below that $\bar{m}$ is identified, and that counterfactuals of m, when x varies while e stays constant, are identified from counterfactuals of $\overline{m}$, when x varies while $\bar{e}$ stays constant. Counterfactuals of m when e varies while x stays constant requires specifying Λ. This is similar to existing identification results for nonseparable models. For example, in the univariate case ($\dim Y = 1$), $m(x, e)$ is identified under A.1-A.2 up to an unknown transformation of ε; see Matzkin (2003). In a similar spirit, the representation results of Carlier, Chernozhukov and Galichon (2016) relies on normalizing the distribution of ε to be known. Here, the normalization is done in terms of the solution mapping Λ since this is the tool used for identification. Since Λ is a functional of $f_{\varepsilon|Z}$, imposing a normalization on Λ implicitly normalizes this density and allows its identification.

Whether $\Lambda(z)$ is a singleton, or a set, or the empty set, is testable, under the maintained Assumptions 1-2. If, in addition to being a singleton, Λ is one–to–one, we will be able to characterize the value of ε by y, rather than by z. In such case, we can identify counterfactuals of the type $m(x', e) - y$, where e is characterized as the value of ε such that $m(x, e) = y$. Whether Λ is one–to–one is also testable under the maintained Assumptions 1-2.

The particular choice of Z determines the properties of Λ. For example, point identification of $\bar{m}$ can only be achieved if $\Lambda(z)$ is a function. Otherwise, we can only partially identify $\bar{m}$. Thus, the researcher should choose Z to ensure maximal covariation between Z and ε in terms of $f_{\varepsilon|Z}(e|z)$. 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.

The identifying power of Λ, as measured by its volume, 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

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

and so the solutions mappings for the score equations of $\bar{\varepsilon}|\bar{Z}$, $\bar{\Lambda}$ satisfies $\bar{\Lambda}(\bar{z}) = G_\varepsilon\left(\Lambda\left(G_Z^{-1}(z)\right)\right)$. Since we will only be able to achieve identification up to the unknown solution mapping, we can without loss of generality always normalize the distributions of of ε and Z. For example, we can normalize them so that their individual components are mutually independent with each component having a pre-specified distribution.

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 2 *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 $\frac{\partial s(z)}{\partial z}$ has full rank and we normalize the distribution of η so that it has a unique mode at zero, then $\Lambda(z) = t^{-1}(s(z))$ is a singleton.

Example 3 *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. As explained above, we can normalize the marginal distributions of ε and Z to be uniform without loss of generality, in which case

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

Taking derivatives with respect to z on both sides of this equation,

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

Thus, the properties of Λ are determined by the ones of $s(\varepsilon, \eta)$ and η. Suppose, for example, that $s(\varepsilon, \eta) = s(\varepsilon) + \eta$ and we normalize the distribution of η so that it has a unique mode at zero; then

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

and so

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

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

The following theorem states our formal identification result:

Theorem 4 *Suppose that Assumptions 1-2 are satisfied. Then the following results hold:*

1. *For any given* $(x, \bar{e}) \in \mathcal{X}_0 \times \mathcal{Z}$, $m(x, \Lambda(\bar{e})) = \bar{m}(x, \bar{e})$ *is (set) identified as the solution(s)* y^* *to*

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

   In particular, if for a given $\bar{e} \in \mathcal{Z}$, *there exists a unique solution* $\Lambda(\bar{e}) \in \mathcal{E}$ *to (4), then* $m(x, \Lambda(\bar{e})) = \bar{m}(x, \bar{e})$ *is point identified as the unique solution* y^* *to (8).*

2. *For any given* $e \in \Lambda(\bar{e})$ *with* $\bar{e} \in \mathcal{Z}$, *and any* $(x, x') \in \mathcal{X}_0 \times \mathcal{X}_0$:

$$m(x', e) - m(x, e) \in \bar{m}(x', \bar{e}) - \bar{m}(x, \bar{e}).$$

3. *If* Λ *is one–to–one, then an equivalent representation of the model is*

$$Y = \bar{m}(X, \bar{\varepsilon}) \Leftrightarrow \bar{r}(X, Y) = \bar{\varepsilon}, \tag{9}$$

   where $\bar{\varepsilon} = \Lambda^{-1}(\varepsilon)$ *and* $\bar{r}(x, y) = \Lambda^{-1}(r(x, y))$. *As a consequence, the heterogeneity distribution* $\bar{\varepsilon}|Z$ *is also identified.*

Proof. Let $\mathcal{Y}^*(x, \bar{e}) \subseteq \mathcal{Y}$ be the set of solutions to eq. (8). Given that $f_{Y|X,Z}$ is identified from data, $\mathcal{Y}^*(x, \bar{e})$ is also identified. We claim that $\mathcal{Y}^*(x, \bar{e}) = \bar{m}(x, \bar{e})$. To this end, recall eq. (6) where, due to m being invertible, $|\partial r(x, y) / (\partial y)| > 0$ and so

$$\frac{\partial f_{Y|X,Z}(y|x, \bar{e})}{\partial z} = 0 \Leftrightarrow \frac{f_{\varepsilon|Z}(r(x, y)|\bar{e})}{\partial z} = 0. \tag{10}$$

Now, take any element $y^* \in \mathcal{Y}^*(x, \bar{e})$. Combining eqs. (8) and (10), we have that $r(x, y^*) \in \Lambda(\bar{e})$. And so $y^* = m(x, r(x, y^*)) \in \bar{m}(x, \bar{e})$ by definition. This shows that $\mathcal{Y}^*(x, \bar{e}) \subseteq \bar{m}(x, \bar{e})$. Reversely, take any element $y^* \in \bar{m}(x, \bar{e})$. By definition, $y^* = m(x, e^*)$ for some $e^* \in \Lambda(\bar{e})$. By invertibility, $e^* = r(x, y^*)$ and applying (10) we conclude that $y^* \in \mathcal{Y}^*(x, \bar{e})$ and so $\bar{m}(x, \bar{e}) \subseteq \mathcal{Y}^*(x, \bar{e})$. The second part of the theorem follows from the first.

In the case where Λ is one-to-one, an observational equivalent representation of the model is given in eq. (9). The identified function $\bar{m}(x,\bar{e})$ is now an invertible function, and the third part of the theorem follows easily. ■

From eq. (6), we see that properties of Λ are embedded in $\partial f_{Y|X,Z}(y|x,z)/(\partial z)$ under Assumptions 1-2. For example, invertibility of Λ is guaranteed by invertibility of $\partial f_{Y|X,Z}(y|x,z)/(\partial z)$ w.r.t. z. Thus, point identification is testable.

The above identification result allows us to trace the responses of each individual in an heterogenous population when the individual unobserved heterogeneity is characterized by an $\bar{e} \in \mathcal{Z}$. However, in may applications, one may instead characterize the individual unobserved heterogeneity by an initial observed heterogeneous response y. Assumption 1 guarantees that for any (y,x), there exists a unique value of ε satisfying $y = m(x,\varepsilon)$. Hence, $\varepsilon = r(y,x)$ is characterized by (y,x). The counterfactual $y' = m(x',\varepsilon)$, on the value of y when the value of x changes to x' while the value of ε stays fixed must satisfy $\varepsilon = r(y,x) = r(y',x')$. Hence, if ε can be characterized by a value of $\bar{e}$ from an initially observed (y,x), the same value of $\bar{e}$ can be used to identify y' when x changes to x'. Neither finding the value of $\bar{e}$ that characterizes ε nor finding the value of y' requires specifying Λ, merely that Λ is one–to–one, which is testable under Assumptions 1-2.

To state this last identification result formally, let us introduce the following generalized version of $\bar{r}$ introduced in eq. (9),

$$\bar{r}(y,x) := \left\{ z \in \mathcal{Z} : \frac{\partial f_{Y|X,Z}(y|x,z)}{\partial z} = 0 \right\}. \tag{11}$$

This is in general a set-valued function with $\bar{r}(y,x) = \varnothing$ if no solution exists. Given (6), $|\partial r(x,y)/(\partial y)| > 0$ and the definition of Λ, it follows that

$$\bar{r}(y,x) = \Lambda^{-1}(r(x,y)),$$

and so $\bar{r}$ is the generalized inverse of $\bar{m}$:

$$y \in \bar{m}(x,\bar{e}) \Leftrightarrow \bar{e} \in \bar{r}(y,x). \tag{12}$$

In the special case where Λ is invertible, $\bar{r}$ is the actual inverse of $\bar{m}$, as introduced in Theorem 4, and characterizes $r(y,x)$. As the next theorem shows, $\bar{r}$ is enough to identify counterfactuals when ε is characterized by y.

Theorem 5 *Suppose that Assumptions 1-2 hold. Then for any $(x,x') \in \mathcal{X}_0 \times \mathcal{X}_0$ and any individual with $\varepsilon = e$ such that $y = m(x,e)$*

$$m(x',e) - m(x,e) \in \bar{m}(x', \bar{r}(y,x)) - \{y\}$$

where $\bar{r}(y,x)$ is defined in eq. (11). As before, in the special case where $\bar{r}(y,x)$ and $\bar{m}(x,\bar{r}(y,x))$ are singletons, point identification is achieved.

Proof. Let $e \in \mathcal{E}$ be the individual's unobserved component which by definition satisfies $y = m(x,e)$. By definition of $\bar{m}$ and $\bar{r}$, $m(x',e) \in \bar{m}(x',\bar{r}(y,x))$ which proves the result. ■

A few remarks are in order: First, we may not be able to track the value of m for all $\varepsilon \in \mathcal{E}$. This only occurs when $\Lambda(\mathcal{Z}) = \mathcal{E}$. Second, a necessary condition for Λ to be one-to-one is that $d_Z = d_\varepsilon$. 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, when $\Lambda(\mathcal{Z}) = \mathcal{E}$, 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 6 *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,

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

Thus, Λ_{d_Z} is a singleton for all choices of $d_Z \geq 1$ and so nothing is gained in terms of identification 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.

3.1.1 Identification of derivatives

Derivatives of the function m with respect to continuously distributed variables can be identified as well in the case when Λ is a function, rather than a correspondence. In the consumer demand case, for example, these are useful for testing integrability conditions or to calculate income and price effects.

To derive an expression when $\bar{m}(x,\bar{e})$ is a function, substitute $y^* = \bar{m}(x,\bar{e})$ into eq. (6), and then take derivatives w.r.t. x on both sides to obtain

$$G(x,\bar{e})\frac{\partial \bar{m}(x,\bar{e})}{\partial x'} + \frac{\partial^2 f_{Y|X,Z}(\bar{m}(x,\bar{e})|x,\bar{e})}{\partial z \partial x'} = 0.$$

where

$$G(x,\bar{e}) = \left.\frac{\partial^2 f_{Y|X,Z}(y|x,\bar{e})}{\partial z \partial y'}\right|_{y=\bar{m}(x,z)} \in \mathbb{R}^{d_Z \times d_Y}. \tag{13}$$

For any weighting matrix $W(x,\bar{e})$ so that

$$H_m(x,\bar{e}) := G(x,\bar{e})' W(x,\bar{e}) G(x,\bar{e}) \in \mathbb{R}^{d_Y \times d_Y} \tag{14}$$

has full rank, we then obtain

$$\frac{\partial \bar{m}(x,\bar{e})}{\partial x'} = -H_m^{-1}(x,\bar{e}) G(x,\bar{e})' W(x,\bar{e}) \frac{\partial^2 f_{Y|X,Z}(\bar{m}(x,\bar{e})|x,\bar{e})}{\partial y \partial x'}, \tag{15}$$

where the right-hand side is identified. This in turn implies that, for any $e = \Lambda(\bar{e})$, $\partial m(x,e)/(\partial x) = \partial \overline{m}(x,\overline{e})/(\partial x)$ is identified. Similarly,

$$\frac{\partial \bar{m}(x,\bar{e})}{\partial \bar{e}'} = -H_m^{-1}(x,\bar{e}) G(x,\bar{e})' W(x,\bar{e}) \frac{\partial^2 f_{Y|X,Z}(\bar{m}(x,\bar{e})|x,\bar{e})}{\partial z \partial z'}, \tag{16}$$

but in this case this does not imply identification of $\partial m(x,e)/\partial e$ since $\partial \bar{m}(x,\bar{e})/(\partial \bar{e}) = [\partial m(x,e)/(\partial e)]_{e=\Lambda(\bar{e})}$ $\partial \Lambda(\bar{e})/(\bar{e})$.

The requirement that the matrix $H_m(x,\bar{e})$ has full rank is quite natural and will hold under great generality given our identification result stating that eq. (6) has a unique solution at $y^* = \bar{m}(x,\bar{e})$. In fact, as we shall see, this rank condition will show up again in the next section when we develop and analyze a GMM-type estimator of $\bar{m}$; there it corresponds to the usual rank condition needed to achieve (local) identification of GMM estimators. In the case where Λ is one-to-one (so that $d_Z = d_\varepsilon$), $W(x,\bar{e})$ is obsolete and we simply require that $G(x,\bar{e})$ has full rank.

When Λ is one-to-one and e is defined as the value of ε satisfying $y = m(x,e)$, for given y and x, an expression of $\partial m(x,e)/\partial x$ can be obtained similarly. The starting equation in this case is

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

By our arguments above, we know that $z^* = \Lambda^{-1}(r(y,x))$. We assume that $\partial^2 f_{Y|X,Z}(y|x,z^*)/\partial z \partial y'$ and $\partial^2 f_{Y|X,Z}(y|x,z^*)/\partial z \partial z'$ are invertible. Taking derivatives with respect to y and with respect to x, we get

$$\frac{\partial^2 f_{Y|X,Z}(y|x,z^*)}{\partial z \partial z'} \frac{\partial \Lambda^{-1}(r(y,x))}{\partial \varepsilon} \frac{\partial r(y,x)}{\partial y} + \frac{\partial^2 f_{Y|X,Z}(y|x,z^*)}{\partial z \partial y'} = 0$$

and

$$\frac{\partial^2 f_{Y|X,Z}(y|x,z^*)}{\partial z \partial z'} \frac{\partial \Lambda^{-1}\left(r\left(y,x\right)\right)}{\partial \varepsilon} \frac{\partial r\left(y,x\right)}{\partial x} + \frac{\partial^2 f_{Y|X,Z}(y|x,z^*)}{\partial z \partial x'} = 0$$

Since

$$\left.\frac{\partial m\left(x,\varepsilon\right)}{\partial x}\right|_{\varepsilon=r(y,x)} = -\left(\frac{\partial r\left(y,x\right)}{\partial y}\right)^{-1}\left(\frac{\partial r\left(y,x\right)}{\partial x}\right)$$

it follows that

$$\left.\frac{\partial m\left(x,\varepsilon\right)}{\partial x}\right|_{\varepsilon=r(y,x)} = -\left(\frac{\partial^2 f_{Y|X,Z}(y|x,z^*)}{\partial z \partial y'}\right)^{-1}\left(\frac{\partial^2 f_{Y|X,Z}(y|x,z^*)}{\partial z \partial x'}\right)$$

Note, in particular, that the derivative does not depend on the unknown transformation Λ.

3.2 Identification with Discrete Z

We next consider the case where Z has 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

$$\Lambda\left(z_1, ..., z_{d_Y+1}\right) = \left\{e \in \mathcal{E} : f_{\varepsilon|Z}\left(e|z_i\right) = f_{\varepsilon|Z}\left(e|z_j\right), \quad 1 \leq i < j \leq d_Y + 1\right\}. \tag{17}$$

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

$$\bar{m}\left(x,\bar{e}\right) = m\left(x,\Lambda\left(\bar{e}\right)\right), \text{ where } \bar{e} \in \bar{\mathcal{E}} := \left\{\bar{e} \in \mathcal{Z}^{d_Y+1} : \bar{e}_i \neq \bar{e}_j \text{ for } i \neq j\right\}, \tag{18}$$

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

$$f_{Y|X,Z}(y^*|x,\bar{e}_i) = f_{Y|X,Z}(y^*|x,\bar{e}_j), \quad 1 \leq i < j \leq d_Y + 1. \tag{19}$$

From eq. (2), this set of equations is equivalent to

$$f_{\varepsilon|Z}\left(r\left(x,y^*\right)|\bar{e}_i\right) = f_{\varepsilon|Z}\left(r\left(x,y^*\right)|\bar{e}_j\right), \quad 1 \leq i < j \leq d_Y + 1, \tag{20}$$

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

Theorem 7 *Under Assumptions 1-2: For any given* $\bar{e} \in \bar{\mathcal{E}}$, $\bar{m}(x, \bar{e})$, *as defined in eq. (18), is identified for all* $x \in \mathcal{X}_0$ *as the (set of) solution(s)* y^* *to eq. (19).*

As in the continuous Z case, the above theorem only allows us to identify $m(x, e)$ at the values of $e \in \Lambda(\bar{e})$ for some $\bar{e} \in \mathcal{Z}^{d_Y+1}$. Thus, given that $\mathcal{Z}$ is countable, we can only identify $m(x, e)$ at a countable number of values $e \in \mathcal{E}$. At the same time, for the consumers that can be reached through (19), we can identify the differences $m(x', e) - m(x, e) = \bar{m}(x, \bar{e}) - \bar{m}(x, \bar{e})$, when the value of x changes to x' while the value of $\bar{e}$ stays fixed, and we can also identify marginal effects, $\partial m(x, e) / (\partial x) = \partial \bar{m}(x, \bar{e}) / (\partial x)$.

4 Nonparametric Estimation and Inference

Let (Y_i, X_i, Z_i), $i = 1, ..., n$, be i.i.d. observations from the model. We wish to use the above identification results to develop nonparametric estimators of the structural function $\bar{m}$. We consider in turn the case of Z being continuously or discretely distributed. To avoid cumbersome notation, we will in the following assume that X has a continuous distribution. We will briefly indicate how the proposed estimators and asymptotic theory have to be adjusted when X contains discrete components. Moreover, in our theoretical analysis, we restrict ourselves to the case where Λ, and therefore $\bar{m}$, is a function. The more general case where $\bar{m}$ is a set-valued function is left for future research.

4.1 Estimation with Continuous Z

Theorem 4 suggests the following GMM-type estimator of $\bar{m}(x, \bar{e}) := m(x, \Lambda(\bar{e}))$ for any given values of $(x, \bar{e}) \in \mathcal{X}_0 \times \bar{\mathcal{E}}$:

$$\hat{m}(x, \bar{e}) = \arg\min_{y \in \mathcal{Y}_0} \hat{g}(y|x, \bar{e})' \hat{W}(x, z)\, \hat{g}(y|x, \bar{e}), \tag{21}$$

where

$$\hat{g}(y|x, \bar{e}) = \frac{\partial \hat{f}_{Y|X,Z}(y|x, \bar{e})}{\partial z'} \tag{22}$$

contain the "moment" conditions, $\hat{f}_{Y|X,Z}$ is a nonparametric estimator of $f_{Y|X,Z}$, $\hat{W}(x, z) \in \mathbb{R}^{d_Z \times d_Z}$ is a weighting matrix, and $\mathcal{Y}_0 \subseteq \mathcal{Y}$ is some compact subset that the true function value $\bar{m}(x, \bar{e})$ lies 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.

In the general case where $\bar{m}$ is a set-valued function, the above estimator $\hat{m}(x, \bar{e})$ will also be set-valued. The analysis of the estimator in this general setting involves more complex arguments and so for simplicity we will in the following focus on the case where $\bar{m}$ is a function, which is implied by the following assumption:

Assumption 3. (i) The solution mapping $\Lambda(z)$ defined in eq. (5) is a function; (ii) $\Lambda(z)$ is one-to-one and so has a well-defined inverse $\Lambda^{-1}(z)$.

If we include Assumption 3(ii), so that Λ is one-to-one and $d_Z = d_Y$, $\hat{g}(y|x,\bar{e})$ just identifies $\bar{m}(x,\bar{e})$ and the weighting matrix $\hat{W}(x,z)$ can be set to the identity matrix. The second part of Theorem 4 also shows that if we add Assumption 3(ii), estimators of the inverse of $\bar{m}(x,\bar{e})$, as denoted $\bar{r}(x,y)$, can be obtained by either computing

$$\hat{r}(x,y) = \arg\min_{\bar{e}\in\bar{\mathcal{E}}_0} \|\hat{g}(y|x,\bar{e})\|^2, \tag{23}$$

or

$$\tilde{r}(x,y) = \arg\max_{\bar{e}\in\bar{\mathcal{E}}_0} \hat{f}_{Y|X,Z}(y|x,\bar{e}), \tag{24}$$

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

$$\hat{f}_{Y|X,Z}(y|x,z) = \frac{\sum_{i=1}^{n} K_{Y,h_Y}(Y_i - y) K_{X,h_X}(X_i - x) K_{Z,h_Z}(Z_i - z)}{\sum_{i=1}^{n} K_{X,h_X}(X_i - x) K_{Z,h_Z}(Z_i - z)}, \tag{25}$$

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 the following restrictions on the kernel functions used to compute $\hat{f}_{Y|X,Z}$, the weighting matrix $\hat{W}_n$, and the underlying structure of the model at the values $(x,\bar{e})$ at which we wish to estimate $\bar{m}$:

Assumption 4. The kernel functions are twice continuously differentiable, of order 2, 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 5. (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 $f_{X,Z}(x,\bar{e}) > 0$.

Assumption 6. (i) $\hat{W}(x,\bar{e}) \to^P W(x,\bar{e}) \in \mathbb{R}^{d_Z \times d_Z}$; (ii) $\bar{m}(x,\bar{e})$ is situated in the interior of $\mathcal{Y}_0$; (iii) $H_m(x,\bar{e})$ defined in eq. (14) has full rank.

Assumptions 4 and 5 allow us to apply standard results from the analysis of nonparametric kernel estimators. In particular, Assumption 5 together with Assumptions 1 and 3 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 is twice continuously differentiable. This combined with the use of second-order kernels, as imposed in Assumption 4, imply that the leading bias terms of $\partial \hat{f}_{Y|X,Z}(y|x,z)/(\partial z')$ are of order $O_P\left(h_Y^2\right)+O_P\left(h_X^2\right)+O_P\left(h_Z^2\right)$, 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 $\hat{m}(x,\bar{e})$ and contains standard conditions found in the analysis of GMM estimators: Assumption 6(i) together with the identification result in Theorem 4 ensure that the GMM estimator defined in eq. (21) is consistent; Assumption 6(ii) rules out that the "true" parameter lies on the boundary of the parameter space; and Assumption 6(iii) is the usual rank condition for GMM estimators that guarantee local identification. If Assumption 3(ii) hold, Assumption 6(i) becomes void and 6(iii) reduces to the requirement that $G(x,\bar{e})$ in eq. (13) has full rank.

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 8 *Suppose that Assumptions 1–2, 3(i), and 4-6 hold. Then, for any bandwidth sequences satisfying*

$$nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} h_a^4 \to 0 \text{ for } a = Y,X,Z,\ \ \log(n)/\left(nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}\right) \to 0,\ \text{ and } nh_Y^{d_Y+2} h_X^{d_X} h_Z^{d_Z+2} \to \infty, \tag{26}$$

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

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

where

$$V_m(x,\bar{e}) = H_m^{-1}(x,\bar{e})\, G(x,\bar{e})'\, W(x,\bar{e})\, \Omega_m(x,\bar{e})\, W(x,\bar{e})\, G(x,\bar{e})'\, H_m^{-1}(x,\bar{e}),$$

and

$$\Omega_m(x,\bar{e}) = \left.\frac{f_{Y|X,Z}(y^*|x,\bar{e})}{f_{X,Z}(x,\bar{e})}\right|_{y^*=\bar{m}(x,\bar{e})} \int_{\mathbb{R}^{d_Y}} K_Y^2(y)\,dy \int_{\mathbb{R}^{d_X}} K_X^2(x)\,dx \int_{\mathbb{R}^{d_Y}} \frac{\partial K_Y(y)}{\partial y}\frac{\partial K_Y(y)}{\partial y'}dy \in \mathbb{R}^{d_Y\times d_Y}. \tag{27}$$

Remark 9 *The first and second bandwidth condition in eq. (26) control the bias and variance of $\hat{g}(y|x,\bar{e})$, respectively, and ensure that they vanish sufficiently fast. The third condition implies that the nonparametric estimator $\hat{G}(x,\bar{e}) = \partial^2 \hat{f}_{Y|X,Z}(y^*|x,\bar{e})/(\partial z \partial y')$ of $G(x,\bar{e})$ 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, $V_m(x,\bar{e})$, 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, $\Omega_m(x,\bar{e})$ 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, $H_m(x,\bar{e})$ measures the identifying strength of Z as it measures the local curvature of the first-order conditions identifying $\bar{m}$. Finally, the usual results regarding efficiency of GMM estimators carry over to our setting with an efficient estimator resulting from choosing $\hat{W}(x,\bar{e})$ as a consistent estimator of $\Omega_m^{-1}(x,\bar{e})$.

Next, we analyze the two estimators of $\bar{r}(x,y)$ defined in eqs. (23)-(24) which can be employed when Assumption 3(ii) also holds. 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 7. (i) $\bar{r}(x,y)$ is situated in the interior of $\bar{\mathcal{E}}_0$ and (ii)

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

Theorem 10 *Suppose that Assumptions 1–7 hold. Then, for any bandwidth sequences satisfying*

$$nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2} h_a^4 \to 0 \text{ for } a = Y,X,Z, \;\; \log n / h_Y^{d_Y} h_X^{d_X} h_{Z|}^{d_Y+4} \to 0, \;\; \text{and } nh_Y^{d_Y+2} h_X^{d_X} h_Z^{d_Z} \to \infty, \tag{28}$$

the estimator $\hat{r}(x,y)$, as defined by eq. (23), is consistent and satisfies

$$\sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2}}\{\hat{r}(x,y) - \bar{r}(x,y)\} \to^d N\left(0, H_r^{-1}(x,y)\,\Omega_r(x,y)\,H_r^{-1}(x,y)\right),$$

where

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

Furthermore, the estimator $\tilde{r}(x,y)$ defined in eq. (24) is first-order equivalent to $\hat{r}(x,y)$.

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 C. As can be seen from those results our estimator performs well in sample sizes around $n = 2,000$. This sample size was chosen to match the one used in the empirical application, and so we expect that the estimator will do well there as well.

4.2 Estimation with Discrete Z

In the discrete case, for any $(x,\bar{e}) \in \mathcal{X}_0 \times \bar{\mathcal{E}}$, $\bar{m}(x,\bar{e})$ is identified as the solution to the restrictions in eq. (19). This 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. (19),

$$\hat{f}_{Y|X,Z}(y|x,z_i) = \hat{f}_{Y|X,Z}(y|x,z_j), \quad 1 \le i < j \le d_Y + 1. \tag{29}$$

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

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

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. (29) as GMM estimator: Let $\hat{g}(y|x,\bar{e}) = \{\hat{g}_{i,j}(y|x,\bar{e}) : 1 \le i < j \le d_Y\}$, where $\hat{g}_{i,j}(y|x,\bar{e}) = \hat{f}_{Y|X,Z}(y|x,\bar{e}_i) - \hat{f}_{Y|X,Z}(y|x,\bar{e}_j)$, contain the "moment restrictions" and define

$$\hat{m}(x,\bar{e}) = \arg\min_{y\in\mathcal{Y}_0} \|\hat{g}(y|x,\bar{e})\|\,. \tag{31}$$

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

For the asymptotic analysis, we maintain Assumption 4, but can dispense of Assumption 5(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, $g(y|x,\bar{e}) = \{g_{i,j}(y|x,\bar{e}) : 1 \le i < j \le d_Y\}$ where $g_{i,j}(y|x,\bar{e}) = f_{Y|X,Z}(y|x,\bar{e}_i) - f_{Y|X,Z}(y|x,\bar{e}_j)$. We then have:

Theorem 11 *Assume that Assumptions 1-3(i), 4 and 5(i) hold and* $H_m(x,\bar{e}) = G(x,\bar{e})' G(x,\bar{e}) \in \mathbb{R}^{d_Y \times d_Y}$ *has full rank where*

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

Then, for any bandwidth sequences satisfying $nh_Y^{d_Y} h_X^{d_X} h_a^4 \to 0$, *for* $a = Y, X$, *and* $nh_Y^{d_Y+2} h_X^{d_X} \to \infty$, $\hat{m}(x,\bar{e})$, *as defined by eq. (31), is consistent and satisfies*

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

where

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

4.3 Counterfactual Inference

As in the statement of Theorem 5, we distinguish between the following two cases: In the first case, we index the population in terms of $\bar{e}$ as defined in terms of the function Λ; in the second case, we index individuals in the population in terms of (Y, X). In either case, we will assume that Assumptions 1-3 are satisfied.

For the first case, we consider a given value $\bar{e}_0 \in \mathcal{Z}$ and employ the stochastic representation given in (9), where $\bar{m}$ and the distribution of $\bar{\varepsilon}$ are identified. Let $\bar{\varepsilon} = \bar{e}_0$; we are interested in measuring how this individual would respond to a change in X from x_0 to x_1: As in Theorem 1, the response is given by

$$\Delta(\bar{e}_0) := \bar{m}(x_1, \bar{e}_0) - \bar{m}(x_0, \bar{e}_0).$$

The resulting estimator then takes the form

$$\hat{\Delta}_Z(\bar{e}_0) := \hat{m}(x_1, \bar{e}_0) - \hat{m}(x_0, \bar{e}_0),$$

where "Z" is used to indicate that the value of ε, which remains fixed as X changes from x_0 to x_1, is characterized by a value of Z. The asymptotic distribution can be derived from

$$\hat{\Delta}_Z(\bar{e}_0) - \Delta(\bar{e}_0) = \{\hat{m}(x_1, \bar{e}_0) - \bar{m}(x_1, \bar{e}_0)\} - \{\hat{m}(x_0, \bar{e}_0) - \bar{m}(x_0, \bar{e}_0)\}, \tag{32}$$

where the two terms on the right-hand side are asymptotically independent of each other by the usual arguments for kernel-based estimators and so again $\hat{\Delta}_Z(\bar{e}_0)$'s large-sample distribution follows directly from Theorem 8.

For the second case, we assume that y_0 and x_0 are given and define e_0 as the value of ε satisfying $y_0 = m(x_0, e_0)$ or alternatively $e_0 = r(y_0, x_0)$. In this case, we first obtain as estimate for e_0: $\hat{e}_0 = \hat{r}(x_0, y_0)$ and then use it to estimate

$$\hat{\Delta}_{y_0}(e_0) := \hat{m}(x_1, \hat{e}_0) - y_0,$$

where "y_0" indicates that the value of ε is characterized by a given value of y. In order to conduct inference using $\hat{\Delta}_{y_0}(e_0)$, we have to take into account the first-step estimation of e_0. This can be done using the delta method,

$$\begin{aligned}\hat{\Delta}_{y_0}(e_0) - \Delta(e_0) &= \hat{m}(x_1, \hat{e}_0) - m(x_1, e_0) \\ &= \frac{\partial \hat{m}(x_1, \tilde{e}_0)}{\partial \bar{e}'}\{\hat{r}(x_0, y_0) - r(x_0, y_0)\} + \{\hat{m}(x_1, e_0) - m(x_1, e_0)\},\end{aligned}$$

where $\tilde{e}_0$ is situated on the line segment connecting $\hat{e}_0$ and e_0; in particular, $\tilde{e}_0 \to^P e_0$ and so $\partial \hat{m}(x_1, \tilde{e}_0)/(\partial e') \to^P \partial \bar{m}(x_1, e_0)/(\partial e')$. Since $\hat{r}(x_0, y_0)$ and $\hat{m}(x_1, e_0)$ are independent in large samples by the usual arguments for kernel estimators, the large sample distribution now follows by combining Theorems 8 and 10. We collect the two results in the following theorem:

Theorem 12 *Suppose that Assumptions 1–6 hold together with eq. (26). Then,*

$$\sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2}}\{\hat{\Delta}_Z(\bar{e}_0) - \Delta(\bar{e}_0)\} \to^d N(0, V_m(x_0, \bar{e}_0) + V_m(x_1, \bar{e}_0)),$$

where $V_m(x, \bar{e})$ is defined in Theorem 8. If furthermore Assumption 7 holds together with eq. (28), then

$$\sqrt{nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2}}\{\hat{\Delta}_{y_0}(e_0) - \Delta(e_0)\} \to^d N(0, V_{y_0}(e_0)),$$

where, with H_r and V_r defined in Theorem 10,

$$V_{y_0}(\bar{e}_0) = \frac{\partial m(x_1, \bar{e}_0)}{\partial \bar{e}'} H_r^{-1}(x_0, y_0) V_r(x_0, y_0) H_r^{-1}(x_0, y_0) \frac{\partial m(x_1, \bar{e}_0)}{\partial \bar{e}'} + V_m(x_1, \bar{e}_0).$$

Remark 13 *The derivative $\partial \bar{m}(x, \bar{e})/(\partial \bar{e})$ can be estimated by replacing population quantities by their nonparametric estimators in eq. (16).*

5 Choosing External Covariates

To achieve point identification we need to identify relevant external covariates Z so that Assumption 3(i) is satisfied. We here focus on the continuous case; the following discussion is easily adapted to the discrete case with obvious modifications.

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 $\hat{m}(x, \bar{e})$ suffers from a curse-of-dimensionality with the precision deteriorating as d_Z increases, c.f. Theorem 14. Second, if $d_Z > d_Y$, the estimator $\hat{m}(x, \bar{e})$ is not invertible in $\bar{e}$ 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.

5.1 Testing for Existence of Sufficient External Covariates

In the following, we take Assumptions 1-2 as maintained hypotheses and then wish to test Assumption 3.1. Consider first the case where we have exactly d_Y external covariates whose validity we wish to test. To that end, first observe that

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

Thus, the rank condition imposed on the matrix $H_r(x,y)$ in Theorem 10 is satisfied if and only if $\partial^2 f_{\bar{\varepsilon}|Z}(\bar{\varepsilon}|\bar{e}^*)/(\partial z \partial z')$ 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 $H_r(x,y)$ is d_Y or not. The matrix $H_r(x,y)$ 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 $\hat{W}(x,\bar{e}) \to^P \Omega_m^{-1}(x,\bar{e})$, where $\Omega_m(x,\bar{e})$ is defined in Theorem 14. It now follows from the limit results derived in the proof of Theorem 14 in conjunction with the arguments in Newey and McFadden (1994, Section 9.5) that

$$\hat{J}(x,\bar{e}) := nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} \min_{y \in \mathcal{Y}_0} \hat{g}(y|x,\bar{e})' \hat{W}(x,\bar{e}) \hat{g}(y|x,\bar{e}) \to^d J(x,\bar{e}) \quad (33)$$

for all $(x, \bar{e})$ under Assumption 3(i), where $J(x, \bar{e}) \sim \chi^2_{d_Z - d_Y}$ and $J(x_1, \bar{e}_1) \perp J(x_2, \bar{e}_2)$ for any two pairs $(x_1, \bar{e}_1)$ and $(x_2, \bar{e}_2)$. Under the alternative, so that Assumption 3(i) does not hold for some $(x, \bar{e})$, we have $\hat{J}(x, \bar{e}) \to^P +\infty$.

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.

5.2 Constructing an index Z

Instead of selecting a subset from the d_Z variables, one can try to to construct an index $Z^{(0)}$ from the set of candidate variables. This can be done in a number of ways. At the most general level, the hypothesis of interest is that

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

A natural way to estimate B is by searching across functions and choose the one that provides the best fit in terms of explaining the variation in Y conditional on X. This can be done by maximum-likelihood methods as described below.

Consider first the case where B is linear so that

$$Z^* = BZ \text{ for some matrix } B \in \mathbb{R}^{d_Y \times d_Z}.$$

The corresponding estimator of $B(z)$ then takes the form $\hat{B}(z) = Bz$ where

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

where

$$\hat{f}_{Y|X,Z^*}(y|x, z^*) = \frac{\sum_{i=1}^{n} K_{Y,h_Y}(Y_i - y) K_{X,h_X}(X_i - x) K_{Z,h_Z}(BZ_i - z^*)}{\sum_{i=1}^{n} K_{X,h_X}(X_i - x) K_{Z,h_Z}(BZ_i - z^*)}. \tag{34}$$

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 $\bar{m}(x, \bar{e})$ as derived earlier.

More generally, 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

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

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 $\hat{f}_{Y|X,Z^*}(Y_i|X_i, B(Z_i))$ is on the form of eq. (34) with BZ_i replaced by $B(Z_i)$. An asymptotic theory for this estimator is outside the scope of this work and is left for future research.

5.3 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 $\bar{m}(x, \bar{e})$. 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 $\hat{G}(y|x, \bar{e}) = (\hat{G}_1(y|x, \bar{e})', ..., \hat{G}_M(y|x, \bar{e})')' \in \mathbb{R}^{Md_Y}$ where

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

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

$$\hat{m}(x, \bar{e}) = \arg\min_{y \in \mathbb{R}^{d_Y}} \hat{G}(y|x, \bar{e})' \hat{W}(x, \bar{e}) \hat{G}(y|x, \bar{e}), \tag{35}$$

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 $G(y|x, \bar{e}) = (G_1(y|x, \bar{e})', ..., G_M(y|x, \bar{e})')' \in \mathbb{R}^{Md_Y}$ where

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

The following theorem generalizes Theorem 8, where, for simplicity, we assume that the same bandwidths is used across the M density estimates:

Theorem 14 *Suppose that Assumptions 1, 2*-3* and 4-5 hold,* $\hat{W}(x,\bar{e}) \to^P W(x,\bar{e})$*, and the matrix*

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

has full rank, where $G_y(x,e) := \partial G(y|x,\varepsilon) / (\partial y')|_{y=\bar{m}(x,\bar{e})} \in \mathbb{R}^{Md_Y \times d_Y}$*. 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* $nh_Y^{d_Y+2} h_X^{d_w} h_Z^{d_Y+2} \to \infty$*,* $\hat{m}(x,\bar{e})$*, as defined by eq. (35), is consistent and satisfies*

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

where

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

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

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

and, for $i \neq j$*,*

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

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 $\hat{W}(x,\bar{e})$ to be a consistent estimator of $W(x,\bar{e}) = V^{-1}(x,\bar{e})$ in which case the asymptotic variance of $\hat{m}(x,e)$ takes the form $\Omega(w,\bar{e}) = \left[Gy(x,\bar{e})' V^{-1}(x,\bar{e}) G_y(x,\bar{e})\right]^{-1}$.

6 Counterfactual Predictions of Individual Consumer Demand

Here we turn our attention to identifying (bounds on) structural demand functions subject to revealed preference inequalities. Our proposed identification and estimation strategy for consumer demands with multidimensional nonseparable heterogeneity can be used for predicting demand counterfactuals for continuous prices or when only finite price variation is available in data. Here we focus on the discrete price case. We also outline how our methods could be used to measure individual welfare counterfactuals.

6.1 Estimation and Inference of Bounds on Counterfactual Demands

Suppose we have observed a repeated cross-section of observed demands and incomes across a finite set of $T \geq 1$ price regimes, $(Y_i(t), I_i(t), Z_i(t))$, $i = 1, ..., n$ and $t = 1, ..., T$, where the number of price regimes T is small or moderate relative to the sample size n in each regime. Here, a new random sample of consumers are collected for each price regime so the samples across different price regimes are mutually independent. In price regime $t \in \{1, ..., T\}$, data is assumed to have been generated by the following relationship

$$Y_i(t) = m(p(t), I_i(t), \varepsilon_i(t)), \quad i = 1, ..., n,$$

so that the observed consumers all face the common price $p(t)$ which we treat as observed. For notational simplicity we suppress any dependence on additional observed covariates, W, which can be thought of as being kept fixed at a particular value as chosen by the researcher in the following. Finally, we require that a given individual's value of ε does not vary across price regimes, and so is meant to reflect preferences that are time-invariant.

Using the techniques developed in the previous section, we can obtain an estimator of $\bar{m}(t, I, e) := m(p(t), I, \bar{e})$ as a function of $(I, \bar{e})$ using data in price regime $t = 1, ..., T$. This can in turn then be employed to construct bounds for counterfactual individual demands for a consumer characterized by a particular value $\bar{e} \in \mathcal{E}$ at existing prices $p(t)$ and income level I_0, or for counterfactual predictions for consumers facing a *new* set of prices $p_0 \notin \{p(t), ..., p(T)\}$ and income level I_0. To do so, we assume that the consumers in the population of interest satisfies the generalized axiom of preferences (GARP), which in turn imposes bounds on demands in counterfactual price regimes, see Afriat (1967) and Varian (1982).

The original implementation of the bounds assume that we have directly observed demands for a given consumer across the T price regimes. This is not the case here, since we only have repeated cross-sections available. To handle this issue, we modify the procedure for estimation of demand counterfactuals proposed in Blundell, Browning and Crawford (2008) to take as input our demand function estimator: First, for $t = 1, ..., T$, with $\hat{m}(t, I, \bar{e})$ denoting our estimator of the demand in price regime t, we estimate the so-called intersection demand levels defined as

$$\hat{y}(t) = \hat{m}(t, \hat{I}(t), \bar{e}), \text{ where } \hat{I}(t) \text{ solves } p_0' \hat{m}(t, \hat{I}(t), \bar{e}) = I_0. \tag{36}$$

However, even if the underlying data-generating demand process satisfies SARP, $\{\hat{y}(t)\}_{t=1}^T$ may not. We therefore proceed as in Blundell et al (2008) and first adjust the first-step estimates to ensure that SARP is satisfied,

$$\{\hat{y}_{\mathcal{C}}(t)\}_{t=1}^{T} = \arg\min_{\{q(t)\}_{t=1}^{T}} \sum_{t=1}^{T} \hat{w}(t)(y(t) - \hat{y}(t))^2 \text{ s.t. } \{y(t)\}_{t=1}^{T} \in \mathcal{C}, \tag{37}$$

where

$$\mathcal{C} = \left\{ \{y(t)\}_{t=1}^{T} : \exists \{V(t), \lambda(t)\}_{t=1}^{T} \text{ so that } \begin{array}{c} V(t) - V(s) + \lambda(t)p(t)'[y(s) - y(t)] \geq 0 \\ \text{and } \lambda(t) \geq 0 \text{ for } s, t = 1, ..., T \end{array} \right\}$$

contains the Afriat inequalities that demand should satisfy under GARP; see Blundell et al (2008) for further details. Finally, we use $\{\hat{y}_{\mathcal{C}}(t)\}_{t=1}^{T}$ to compute the the following support set estimator containing the set of predicted demands for the particular consumer,

$$\hat{\mathcal{S}}_{p_0,I_0,e} = \left\{ y_0 \in \mathbb{R}^{d_Y} : \begin{array}{c} y_0 \geq 0, \quad p_0'y_0 = I_0, \text{ and} \\ p(t)'y_0 \geq p(t)'\hat{y}_{\mathcal{C}}(t) \text{ for } t = 1, ..., T \end{array} \right\}. \tag{38}$$

Blundell, Kristensen and Matzkin (2014) provides an asymptotic theory for this estimated support set in the case where the estimated demand function has been restricted to satisfy GARP (and so $\hat{y}_{\mathcal{C}}(t) = \hat{y}(t)$). We do not impose this condition in the estimation of m, instead we first adjust the initial demand estimates to obtain $\{\hat{y}_{\mathcal{C}}(t)\}_{t=1}^{T}$ which we then use to estimate the support set. The asymptotic theory of Blundell, Kristensen and Matzkin (2014) is easily modified to take this into account, and so consistency and convergence rate of $\hat{\mathcal{S}}_{p_0,x_0,\varepsilon}$ can be established applying the arguments found there. So we here only focus on how to construct a valid confidence set for $\hat{\mathcal{S}}_{p_0,x_0,\varepsilon}$. To this end, we first derive the asymptotic properties of $\{\hat{y}(t)\}_{t=1}^{T}$, which can be done by formulating $\left\{\hat{I}(t)\right\}_{t=1}^{T}$ as a Minimum Distance Estimator: For given values of p_0, I_0 and e, $\hat{I}(t)$ can be expressed as

$$\hat{I}(t) = \arg\min_{I \in \mathcal{I}_0} \left\| \hat{G}(t, I) \right\|, \quad \hat{G}(t, I) = p_0'\hat{m}(t, I, \bar{e}) - I_0,$$

for some compact set $\mathcal{I}_0 \subset \mathbb{R}_+$.

Applying standard arguments for the analysis of GMM Estimators, we then obtain the asymptotic distribution of $\hat{I}(t)$ which combined with the delta method provides us with the asymptotic distribution of $\hat{y}(t) = \hat{m}(t, \hat{I}(t), \bar{e})$, c.f. Lemma 17,

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

where, with $\bar{I}(t)$ solving $p_0'm(t, \bar{I}(t), \varepsilon) = I_0$ and $\bar{y}(t) = \bar{m}(t, \bar{I}(t), \bar{e})$,

$$\Sigma(t) = \frac{\partial m(t, \bar{I}(t), \bar{e})}{\partial I} \frac{\partial m(t, \bar{I}(t), \bar{e})'}{\partial I} \left(p_0' \frac{\partial m(t, \bar{I}(t), \bar{e})}{\partial I} \right)^{-1} p_0'\Omega_m\left(t, \bar{I}(t), \bar{e}\right) p_0. \tag{39}$$

Here, $\Omega_m(t, I, \bar{e})$ is the large-sample variance matrix of $\hat{m}(t, I, \bar{e})$, as given in eq. (27) with $x = I$.

With the above result, we can obtain a $1-\alpha$ confidence set for $\mathcal{S}_{p_0,I_0,e}$ by the following argument: Let $Q_{\mathcal{S}}\left(\{y(t)\}_{t=1}^T\right)$ be the solution mapping defined as

$$Q_{\mathcal{S}}\left(\{\hat{y}(t)\}_{t=1}^T\right) = \arg\min_{\{y(t)\}_{t=1}^T} \sum_{t=1}^T \hat{w}(t)\left(\hat{y}(t) - y(t)\right)^2 \text{ s.t. } \{y(t)\}_{t=1}^T \in \mathcal{C},$$

and let $Q_{\text{GARP}}\left(\{y(t)\}_{t=1}^T\right)$ be the set mapping taking any given sequence $\{y(t)\}_{t=1}^T$ satisfying GARP into the set

$$Q_{\text{GARP}}\left(\{y(t)\}_{t=1}^T\right) := \left\{ y_0 \in \mathbb{R}^{d_Y} : \begin{array}{c} y_0 \geq 0, \quad p_0' y_0 = I_0 \\ p(t)' y_0 \geq p(t)' y(t) \text{ for } t = 1, ..., T \end{array} \right\}.$$

With these definitions, we can write our estimated support set as

$$\hat{\mathcal{S}}_{p_0,I_0,e} = Q_{\text{SARP}}\left(Q_{\mathcal{S}}\left(\{\hat{y}(t)\}_{t=1}^T\right)\right). \tag{40}$$

In particular, for a given set $CI_{1-\alpha}$ so that $P\left(\{\hat{y}(t)\}_{t=1}^T \in CI_{1-\alpha}\right) \to 1-\alpha$ as $n \to \infty$, we have

$$P\left(\hat{\mathcal{S}}_{p_0,I_0,e} \in Q_{\text{SARP}}\left(Q_{\mathcal{S}}\left(CI_{1-\alpha}\right)\right)\right) = P\left(\{\hat{y}(t)\}_{t=1}^T \in CI_{1-\alpha}\right) \to 1-\alpha \text{ as } n \to \infty. \tag{41}$$

That is, we can construct a confidence set for $\mathcal{S}_{p_0,I_0,e}$ by first constructing one for $\{\bar{y}(t)\}_{t=1}^T$ as defined in the above lemma, and then map this into the corresponding union of support sets.

One simply way of computing an approximate version of the confidence set is through the following simulation-based procedure:

S.1 Compute $y^*(t) = \hat{y}(t) + e^*(t)$, where $e^*(t)$ is drawn from the uniform distribution on the set

$$\left\{e : e'\hat{\Sigma}^{-1}(t)\, e \leq \chi^2_{d_Y}(1-\alpha) / \left(n h_Y^{d_Y} h_X^{d_X} h_Z^{d_Y+2}\right)\right\},$$

and $\hat{\Sigma}(t)$ is an estimator of $\Sigma(t)$ defined in eq. (39), $t = 1, ..., T$.

S.2 Compute $\{\bar{y}^*_{\mathcal{S}}(t)\}_{t=1}^T = Q_{\mathcal{S}}\left(\{y^*(t)\}_{t=1}^T\right)$.

S.3 Compute $\mathcal{S}^*_{p_0,x_0,\varepsilon} = Q_{\text{SARP}}\left(\{\bar{y}^*_{\mathcal{S}}(t)\}_{t=1}^T\right)$.

Repeating 1.-3, $N \geq 1$ times will provide us with a simulated version of $Q_{\text{SARP}}\left(Q_{\mathcal{S}}\left(CI_{1-\alpha}\right)\right)$. The asymptotic validity of the proposed procedure is stated in the following theorem:

Theorem 15 *Under the Assumptions of Theorem 8,*

$$\left\{\mathcal{S}^*_{p_0,x_0,\varepsilon}(k) : k = 1, ..., N\right\} \to^{P^*} Q_{SARP}\left(Q_{\mathcal{S}}\left(CI_{1-\alpha}\right)\right) \text{ as } N \to \infty,$$

where $\mathcal{S}^*_{p_0,x_0,\varepsilon}(k)$, $k = 1, ..., N$, are independent copies obtained from the simulation-based algorithm S.1-S.3 above, and

$$P\left(\hat{\mathcal{S}}_{p_0,I_0,e} \in Q_{SARP}\left(Q_{\mathcal{S}}\left(CI_{1-\alpha}\right)\right)\right) \to 1-\alpha \text{ as } n \to \infty.$$

6.2 Estimating Bounds on Welfare Counterfactuals

The identification of individual heterogeneous demands developed in this paper allows the recovery of bounds on individual welfare measures. Afriat (1977) showed how revealed preference restrictions can be used to provide information on the curvature of indifference surfaces in commodity space and then used to set bounds on the welfare effects of a price change. This is further developed in Varian (1982) and Manser and McDonald (1988). One problem with applying this procedure to the aggregate data that the latter use is that budget surfaces rarely cross so that the bounds from such data tend to be wide. Knowledge of individual nonparametric expansion paths can greatly improve these bounds on true cost of living indices.

Without loss of generality, consider an indifference surface passing through some base vector of goods y_1 at base price vector p_1 and income I. If GARP and weak normality hold, then Blundell, Browning and Crawford (2003) show that each expansion path vector $y_t(I)$, for price vector p_t, can be partitioned into three distinct segments. First, on any expansion path, there are the demands that can be shown to be weakly revealed preferred to y_1, $Q_B(y_1)$. Second, we have the demands that we can show are weakly revealed dominated by y_1, $Q_W(y_1)$. Finally there is an intermediate segment with demands that cannot be revealed preference ordered with respect to y_1.

These segments for each expansion path allow us to construct tight bounds on the welfare costs of arbitrary price changes from the base price p_1. For example, suppose that we have a reference commodity level y_1 (on the expansion path $y_1(I)$) and an arbitrary price vector p_s. The true cost-of-living index based at y_1 is given by $c(p_s, y_1)/c(p_1, y_1)$ where $c(p_s, y_1)$ is the expenditure function giving the cost of attaining a bundle indifferent to y_1 at prices p_s, and $c(p_1, y_1) = I$. Bounds can be placed on this index using

$$\min_y \{p_s' y | y \in Q_w(y_1)\} \leq c(p_s, y_1) \leq \min_y \{p_s' y | y \in Q_B(y_1)\}.$$

Using the SARP constrained individual expansion paths from the previous subsection to construct the best sets, $Q_B(y_1)$, and worse sets, $Q_w(y_1)$, bounds on the true cost of living index for any change in relative prices can be computed for each individual.

7 Empirical Application

In this section, we use the framework developed in section 6 for the case of the heterogeneous consumer demand models to estimate bounds on individual consumer demand counterfactuals. We use the history of the British Family Expenditure Survey (FES) which provides consumer expenditure data at the household level for a large representative sample of consumers in Britain on a consistent basis over many years. To maintain comparison with earlier work, we choose a similar sample and choice of goods to that in Blundell, Browning and Crawford (2008).

7.1 Data

We use a sample from the FES for the period 1997 to 2006. There are approximately 6000 observation in each wave of the survey. The FES contains detailed expenditure data on families together with income and demographic characteristics. In this analysis we take three broad consumption goods - food, services (serv) and other nondurables, for the waves 1997 to 2006. We select a broad range of households with differing demographic characteristics. We choose nondurables as numeraire and estimate demand for food and services so that $d_Y = 2$. The means of the variables are reported in Table 8.1.[3]

These three consumption goods are of particular interest. The price responsiveness of food relative to services and to other nondurables at different income levels is a key parameter in the indirect tax policy debate. Demand counterfactuals across the population of consumers are a main input into the policy analysis. Although food is largely free of value added tax (VAT) in the United Kingdom, the discussions over the harmonization of indirect tax rates across Europe, the large variations in VAT on food and the implications of a flat uniform expenditure tax across all consumption items requires a clear understanding of individual demand responses to nonmarginal changes in the price of food across the income distribution.

Table 8.1: Summary Statistics

year	ln(income)	share-food	share-serv	price-food	price-serv	family size
1997	5.412	.208	.212	2.84	3.88	2.26
1998	5.477	.206	.214	2.89	4.04	2.26
1999	5.498	.204	.218	2.93	4.21	2.22
2000	5.561	.197	.229	2.96	4.41	2.19
2001	5.601	.194	.233	3.06	4.60	2.23
2002	5.611	.194	.241	3.11	4.84	2.20
2003	5.648	.192	.238	3.17	5.08	2.22
2004	5.667	.193	.239	3.21	5.23	2.18
2005	5.710	.188	.246	3.26	5.37	2.18
2006	5.755	.186	.244	3.34	5.53	2.18

[3] The data are from public use files. Further details are available from the euthors on request.

7.2 Estimated Bounds for Counterfactual Individual Demands on Unobserved Budgets

There are three steps to the empirical analysis. The first is to choose observable characteristics that are likely to covary with unobserved individual heterogeneity in demand. Second, to estimate the unrestricted demands in each price regime - that is an Engel curve (expansion path) for each wave of the data and each individual. Finally, to impose the revealed preference inequalities on individual demands and construct bounds on predicted demands at previously unobserved prices and income.

Observed heterogeneity is represented in two *excluded covariate* indices, Z_1 and Z_2. These covariates are chosen to contain independent variation and reflect heterogeneity in both food and services demand. The first of these, Z_1, was chosen to represent family composition and is a measure of the number of equivalent adults in the family (one for the first adult, .6 for additional adults and .4 for each child). Although important in all consumption decisions, the number of equivalent adults is likely to be directly related to the heterogeneity in food consumption. The second, Z_2, was chosen to represent the generation of the head of household in the family. It is a measure for the year of birth of the head of household adjusted for his or her age at leaving full time education (year of birth of cohort plus age at leaving education). This second index we consider to be more related to the consumption of services. Both of these indices were found to be important independent determinants of the demand for food and services in a standard parametric "quadratic almost ideal" demand system (QUAIDS). The family composition variable Z_1 showing clearly a strong driver of food preferences while the generation variable Z_2 was a strong driver of service demand.[4]

Using Z_1 and Z_2, we obtain an estimator of the demand equations $\bar{m}(t, I, \bar{e})$ at each price regime $p(t)$ as a function of income and unobserved heterogeneity (I, e) as specified in Section 4.1. To estimate the bounds on counterfactual demands for new budgets, we can follow the steps outlined in Section 6.1: For a given consumer characterized by a particular value of unobserved heterogeneity, which in turn corresponds to a particular value of the indices, z_0, we compute demand at a particular income level, I_0, when facing prices p_0.

We have two "inside" goods: food and services, and the "outside" good, which we denote "other"; let $p_0 = (p_{0,\text{food}}, p_{0,\text{serv}}, p_{0,\text{other}})'$ be the counterfactual (unobserved) prices. We choose the income I_0 to be the mean income in 2002, roughly the middle of our sample. We estimated demand shares for food and services as function of log-income with $\bar{e}$ chosen as $\bar{e} = (\bar{e}_1, \bar{e}_2) = (2, 1970)$. We calculated these estimates by using an "adaptive" grid, which becomes denser as it approximates the solution. We chose the bandwidths h_a, $a = Y, X, Z$, by first computing the multivariate version of Silverman's

[4] Results available from the authors on request.

"Rule-of-Thumb" for the estimation of kernel density and its derivatives.

The resulting estimator provides an estimated expansion (expenditure share as a function of total budget I) path by year for the consumer characterized by the chosen value $\bar{e} \in \bar{\mathcal{E}}$. After computing the estimated demand functions at each of the observed price regimes, $p(1), ..., p(T)$, we solved eq. (36) using Matlab's nonlinear equation solver to obtain $\hat{I}(t)$ and thereby $\hat{y}(t) = \hat{m}(t, \hat{I}(t), \bar{e})$, $t = 1, ..., T$. Next, we adjusted these T first-step estimates of intersection demands to ensure that they jointly satisfied SARP; this was done by solving the nonlinear constrained programme given in eq. (37). Finally, we plugged the SARP constrained demand estimates into eq. (38) and computed the estimated set of possible demands, $\hat{\mathcal{S}}_{p_0,I_0,e}$.

In general, this support set for counterfactual demands cannot be represented graphically since its of dimension $d_Y \geq 2$. However, one can compute the upper and lower bounds of this set for a given good $k \in \{1, ..., d_Y\}$ by

$$\begin{aligned}
\hat{y}_{0,k}^{\text{up}} &= \arg\max_{y_k} y_{0,k} \text{ s.t. } y_0 = (y_{0,1}, ., .., y_{0,d_Y}) \in \hat{\mathcal{S}}_{p_0,I_0,e}, \text{ and} \\
\hat{y}_{0,k}^{\text{low}} &= \arg\min_{y_k} y_k \text{ s.t. } y_0 = (y_{0,1}, ., .., y_{0,d_Y}) \in \hat{\mathcal{S}}_{p_0,I_0,e},
\end{aligned}$$

respectively. These are linear programmes and so can be solved using standard numerical solvers, yielding an interval $[\hat{y}_{0,k}^{\text{low}}, \hat{y}_{0,k}^{\text{up}}]$ representing the demand bounds for good $k \in \{1, ..., d_Y\}$ of a consumer with income level I_0 and unobserved heterogeneity e facing price level p_0.

A given demand interval $[\hat{y}_{0,k}^{\text{low}}, \hat{y}_{0,k}^{\text{up}}]$ is characterized by a given choice of p_0, I_0 and $\bar{e}_0$. By varying either prices, p_0, income, I_0, or $\bar{e}_0$, this allows us to estimate bounds for the responses in demand to changes in either price, income, or level of heterogeneity.

Figures 1 and 2 present the estimated demand bounds for food together with their corresponding confidence intervals (95%) when either its own price, $p_{0,\text{food}}$, or the price of services, $p_{0,\text{serv}}$, changes, while keeping all other prices fixed at the price level at time T, and income at the chosen I_0.

For the particular consumer characterized by $(I_0, \bar{e}_0)$, Figure 1 describes how the demand for food responds to counterfactual changes in the price of food, $p_{0,\text{food}}$, keeping all other prices fixed. Figure 2 presents the cross-effect on demand for food when the price of services, $p_{0,\text{serv}}$, varies. The estimated bounds suggest a mildly downward sloping own demand for food demands and a upward sloping cross-demand curve for food with respect to the price of services. The figures show that the bounds on counterfactual demands can be quite tight, especially where the data is dense, becoming wider where there is sparse data.

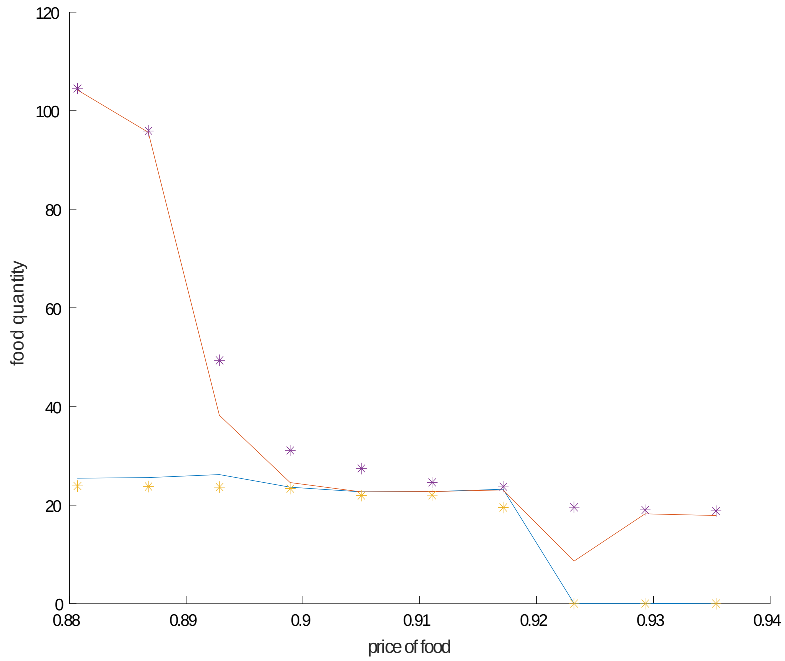

Figure 1: Counterfactual Bounds and CIs on the Demand for Food

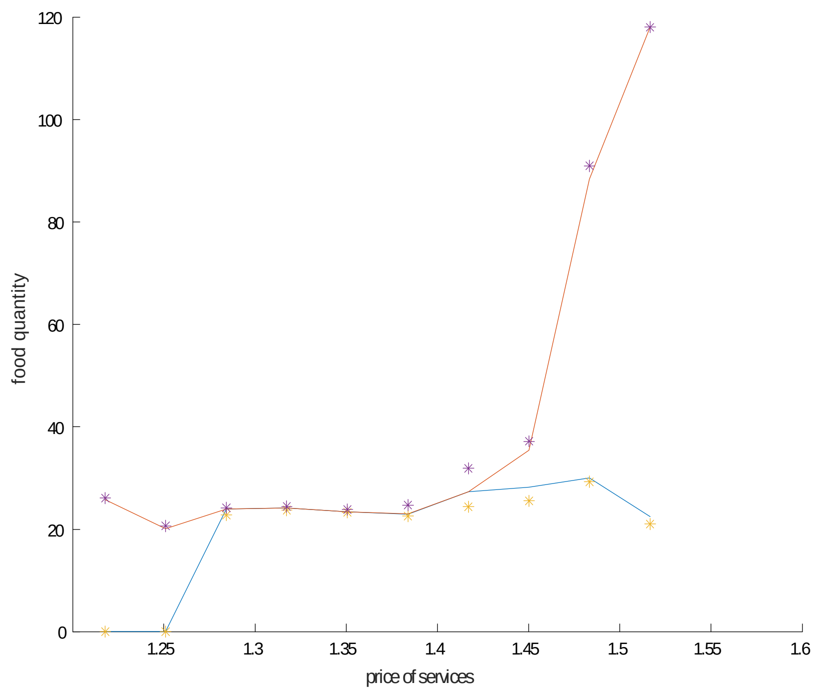

Figure 2: Counterfactual Bounds and CIs on the Cross Demand for Food

8 Summary and Conclusions

We have developed nonparametric methods to identify and estimate counterfactuals for individuals, when each is characterized by a vector of unobserved characteristics. The method requires no functional restrictions on the manner in which the unobservable characteristics affect individual behavior. The general framework corresponds to the reduced form of a nonparametric system of simultaneous equations, where each function depends in unrestricted ways on a vector of unobserved variables.

Our results have relied on two fundamental assumptions: First, the system is invertible in the vector of unobserved heterogeneity. Second, there exist external, individual-specific, covariates that are stochastically related to the unobserved heterogeneity. The external variables do not enter directly into the functions of interest and are related to the unobserved heterogeneity through a nonparametric testable restriction. Once one conditions on the external covariates, the vector of unobserved heterogeneity is independent of the observable covariates appearing in the system, the "internal covariates".

We have applied the methods to consumer choice, where the system of demand functions for each individual is a nonparametric and nonadditive function of prices, income, and a vector of unobserved tastes. In this context, we have demonstrated how the method can be used to estimate individual demand functions from the distribution of demands when individuals are heterogenous in unrestricted ways. We have also described how the demands estimated using our method, together with Revealed Preference inequalities, can be employed to obtain bounds on a counterfactual individual demand for prices and incomes that have not been observed. They can also be used to bound measures of consumer welfare.

The usefulness of the estimators has been illustrated through an empirical application using UK household consumer data. We constructed estimated revealed preference bounds on counterfactual demands for food, services and other goods for each consumer characterized by a given choice of total budget and unobserved heterogeneity. The estimated bounds suggested mildly downward sloping own demand for food and a upward sloping cross-demand curve for food with respect to the price of services. The results have shown that estimated bounds on counterfactual demands can be quite tight, especially where the distribution of observed relative prices is dense, becoming wider where there is sparse price data.

References

Afriat, S. (1967) The Construction of a Utility Function from Demand Data, *International Economic Review* 8, 67-77.

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

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.

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

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

Berry, S. and P. Haile (2015) Identification of Nonparametric Simultaneous Equations Models with a Residual Index Structure, manuscript, Yale University.

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. 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., M. Browning, L. Cherchye, I. Crawford, B. de Rock, and F. Vermeulen (2015) Sharp for SARP: Nonparametric Bounds on the Behavioural and Welfare Effects of Price Changes, *American Economic Journal: Microeconomics* 7, 43–60.

Blundell, R., M. Browning, and I. Crawford (2003) Nonparametric Engel Curves and Revealed Preference, *Econometrica*, 71, 205–240.

Blundell, R., M. Browning, and I. Crawford (2008) Best Nonparametric Bounds on Demand Responses, *Econometrica*, 76, 1227–1262.

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

Brown, D.J., R. Deb, and M.H. Wegkamp (2006) Tests of Independence in Separable Econometric Models: Theory and Application, unpublished manuscript.

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

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.

Cherchye, L., Crawford, I., De Rock, B., and F. Vermeulen (2009) The Revealed Preference Approach to Demand, in *Quantifying Consumer Preferences* (Ed. D. Slottje). Emerald Books.

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

Fan, J., L. Peng, Q. Yao and W. Zhang (2009) Approximating Conditional Density Functions Using Dimension Reduction, *Acta Mathematicae Applicatae Sinica* 25, 445–456.

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

Hansen, B. (2008) Uniform Convergence Rates for Kernel Estimation with Dependent Data, *Econometric Theory* 24, 726–748.

Hausman, J.A. and W.K. Newey (2016) Individual Heterogeneity and Average Welfare, *Econometrica* 84, 1225–1248

Hausman, J.A. and W.K. Newey (2017) Nonparametric Welfare Analysis, *Annual Review of Economics* 9, 521–546.

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.

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 (2013) Nonparametric Analysis of Random Utility Models: Testing, cemmap Working Paper, CWP 36/13, July.

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

Li, K. and J.S. Racine (2006) *Nonparametric Econometrics: Theory and Practice,* Princeton University Press.

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

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

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.

Varian, H. (1982) The Nonparametric Approach to Demand Analysis, *Econometrica*, 50, 945–972.

Appendices

A Proofs

Proof of Theorem 8. 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 $(x,\bar{e}) \in X_0 \times Z$,

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

where $\hat{g}(y|x,\bar{e})$ is defined in eq. (22), and $g(y|x,\bar{e}) = \partial f_{Y|X,Z}(y|x,\bar{e})/(\partial z)$; see, for example, Hansen (2008, Theorem 7). This combined with Assumption 6(i) and the bandwidth restrictions stated in the theorem yield

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

Consistency now follows from Newey and McFadden (1994, Theorem 2.6), where identification is achieved through Theorem 4.

Next, we derive the asymptotic distribution of $\hat{m}(x,\bar{e})$: With $\hat{y}^* := \hat{m}(x,\bar{e})$, $y^* := \bar{m}(x,\bar{e})$ 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

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

where $\hat{G}(x,\bar{e}) = \partial^2\hat{f}_{Y|X,Z}(y^*|x,\bar{e})/(\partial z\partial y')$, $\hat{H}_m(x,\bar{e}) := \hat{G}(x,\bar{e})'\,\hat{W}(x,\bar{e})\,\tilde{G}(x,\bar{e})$, and $\tilde{G}(x,\bar{e}) = \partial^2\hat{f}_{Y|X,Z}(\tilde{y}|x,\bar{e})/(\partial z\partial y')$. Under the stated conditions on the bandwidths in eq. (26), it follows from Lemma 16 that

$$\sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}}\,\{\hat{g}(y^*|x,\bar{e}) - g(y^*|x,\bar{e})\} = \sqrt{nh_Y^{d_Y}h_X^{d_X}h_Z^{d_Z+2}}\hat{g}(y^*|x,\bar{e}) \to^d N(0, \Omega_m(x,\bar{e})),$$

while $\hat{G}(x,\bar{e})$ and $\tilde{G}(x,\bar{e})$ both converge towards $G(x,\bar{e})$ 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 8, and so we only sketch the proof for $\hat{r}(x,y)$. With $\hat{g}(y|x,\bar{e})$ defined in eq. (22) and $g(y|x,\bar{e}) = \partial f_{Y|X,Z}(y|x,\bar{e})/(\partial z)$, we have, for any given $(x,y) \in X \times Y$,

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

This together with the identification result of Theorem 4 shows consistency. To obtain the asymptotic distribution, first observe that, with $\hat{e}^* := \hat{r}(x,y)$ and $\bar{e}^* := r(x,y)$,

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

where $\hat{H}_r(x,y) = \partial^2 \hat{f}_{Y|X,Z}(y|x,\tilde{e})/(\partial z \partial z')$, and $\tilde{e}$ is situated on the line segment connecting $\hat{e}^*$ and $\bar{e}^*$. Under the bandwidth conditions, Lemma 16 implies that

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

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

Proof of Theorem 11. The proof follows along the same lines as the one of Theorem 8 and so is left out. ■

Proof of Theorem 14. The proof follows along the same lines as the one of Theorem 8 and so is left out. ■

Proof of Theorem 15. From Lemma 17, we see that $P(\{\hat{y}(t)\}_{t=1}^T \in CI_{1-\alpha}) \to 1-\alpha$ as $n \to \infty$, where $CI_{1-\alpha} = CI_{1-\alpha}(1) \times \cdots CI_{1-\alpha}(T)$ and

$$CI_{1-\alpha}(t) = \left\{ y : (y - \hat{y}(t))' \hat{\Sigma}^{-1}(t)(y - \hat{y}(t)) \leq \chi^2_{d_Y}(1-\alpha) r_n \right\}.$$

It now follows that $\hat{S}_{p_0,I_0,e}$ defined in eq. (40) satisfies eq. (41).

Consider now the simulated version of $Q_{\text{SARP}}(Q_{\mathcal{S}}(CI_{1-\alpha}))$ defined in the theorem. By construction, $y_k^*(t) = \hat{y}(t) + e_k^*(t)$, where $e_k^*(t)$ is uniformly distributed on the set $\left\{ e : e' \hat{\Sigma}^{-1}(t) e \leq \chi^2_{d_Y}(1-\alpha) r_n \right\}$, satisfies $y_k^*(t) \in CI_{1-\alpha}(t)$. Moreover, as $N \to \infty$, $CI_{k,1-\alpha}^* := \left\{ \{y_k^*(t)\}_{t=1}^T : k = 1, ..., N \right\} \to^{P^*} CI_{1-\alpha}(t)$. Thus, with $S_{p_0,x_0,\varepsilon}^*(k) = Q_{\text{SARP}}\left(Q_{\mathcal{S}}\left(CI_{k,1-\alpha}^*\right)\right)$, the above claim follows by the continuous mapping theorem. ■

B Lemmas

Lemma 16 *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 $nh_Y^{d_Y} h_X^{d_X} h_Z^{d_Z+2} \to \infty$,*

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

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,

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

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

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

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

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

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

Lemma 17 *Assume that, for any given $t = 1, ..., T$:*

(i) $\sqrt{r_n}\{\hat{m}(t, I, e) - m(t, I, e)\} \to^d N(0, \Omega(t, I))$ *for any* $I \in I_0$ *for some* $r_n \to \infty$.

(ii) $\sup_{x \in \mathcal{I}_0} \|\hat{m}(t, I, e) - m(t, I, e)\| \to^P 0$ *and* $\sup_{x \in \mathcal{I}_0} \|\partial \hat{m}(t, I, e)/(\partial I) - \partial m(t, I, e)/(\partial I)\| \to^P 0$.

(iii) $I \mapsto m(t, I, \varepsilon)$ *is strictly increasing.*

Then, the estimated intersection demand $\hat{y}(t)$ defined in eq. (36) satisfies:

$$\sqrt{r_n}\{\hat{y}(t) - \bar{y}(t)\} \to^d N(0, \Sigma(t)),$$

where, with $\bar{I}(t)$ solving $p_0' m(t, \bar{I}(t), \varepsilon) = x_0$ and $\bar{y}(t) = d(t, \bar{I}(t), \varepsilon)$,

$$\Sigma(t) = \frac{\partial m(t, \bar{I}(t), \varepsilon)}{\partial x} \frac{\partial m(t, \bar{I}(t), \varepsilon)'}{\partial x} \left(p_0' \frac{\partial m(t, \bar{I}(t), \varepsilon)}{\partial x}\right)^{-1} p_0' \Omega\left(t, \bar{I}(t)\right) p_0.$$

Proof. Define $G(t, I) = p_0' d(t, I, e) - I_0$. Due to Condition (ii) of the lemma, we have that

$$\sup_{I \in \mathcal{I}_0} \left|\hat{G}(t, I) - G(t, I)\right| \leq \|p_0\| \sup_{I \in \mathcal{I}_0} \|\hat{m}(t, I, e) - m(t, I, e)\| \to^P 0,$$

while Condition (iii) implies that $\bar{I}(t)$ is the unique solution to $G(t, I) = 0$. It now follows from, e.g., Newey and McFadden (1994, Theorem 2.1) that $\hat{I}(t) \to^P \bar{I}(t)$. Next, by (i) and (ii), we have

$$\sqrt{r_n}\left\{\hat{G}\left(t, \bar{I}(t)\right) - G\left(t, \bar{I}(t)\right)\right\} = p_0'\left\{\hat{m}(t, \bar{I}(t), \varepsilon) - m(t, \bar{I}(t), \varepsilon)\right\} \to^d N\left(0, p_0' \Omega\left(t, \bar{I}(t)\right) p_0\right),$$

and, uniformly in I,

$$\frac{\partial \hat{G}(t, I)}{\partial I} - \frac{\partial G(t, I)}{\partial I} = p_0' \left\{\frac{\partial \hat{m}(t, I, e)}{\partial I} - \frac{\partial m(t, I, e)}{\partial I}\right\} \to^P 0.$$

It now follows from Newey and McFadden (1994, Theorem 3.2) that $\sqrt{r_n}(\hat{I}(t) - \bar{I}(t)) \to^d N(0, \Sigma_I(t))$ where

$$\Sigma_I(t) = \left(p_0' \frac{\partial m(t, \bar{I}(t), e)}{\partial I}\right)^{-1} p_0' \Omega\left(t, \bar{I}(t)\right) p_0.$$

By the mean-value theorem together with (ii), this in turn implies

$$\sqrt{r_n}(\hat{y}(t) - \bar{y}(t)) = \left[\frac{\partial m(t, \bar{I}(t), \varepsilon)}{\partial x} + o_P(1)\right] \sqrt{r_n}(\hat{I}(t) - \bar{I}(t)) \to^d N(0, \Sigma(t)),$$

where

$$\Sigma(t) = \frac{\partial m(t, \bar{I}(t), e)}{\partial I} \frac{\partial m(t, \bar{I}(t), e)'}{\partial I} \Sigma_x(t).$$

■

C Simulation Study

Here we investigate the performance of the estimator through simulations. 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 $\eta \sim N\left(\mu_\eta, \Omega_\eta\right)$. Thus, $\varepsilon|Z \sim N\left(Z + \mu_\eta, \Omega_\eta\right)$ and $Y|\left(X, Z\right) \sim N\left(ZX + \mu_\eta, x^2\Omega_\eta\right)$. As such its density is given by

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

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

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

which is the inverse $r\left(x, y\right) = y/x$ of the structural relation $Y = m\left(X, \varepsilon\right) = X\varepsilon$. For given values of (y, x), we implement the estimator of $r(x, y)$ defined as $\hat{r}(x, y) = \arg\max_z \hat{f}_{Y|X,Z}\left(y|x, z\right)$ where $\hat{f}_{Y|X,Z}\left(y|x, z\right)$ 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 $\hat{r}\left(x, y\right) = \left(\hat{r}_1\left(x, y\right), \hat{r}_2\left(x, y\right)\right)$ are reported in Figures 3-6. In each figure we fix q at a particular value, say, $\bar{q}$, and then plot the estimates of the function $x \mapsto r_1\left(x, \bar{q}\right)$ and $x \mapsto r_2\left(x, \bar{q}\right)$. The results show that the kernel-based estimator works quite well, with small biases and not too big variances.

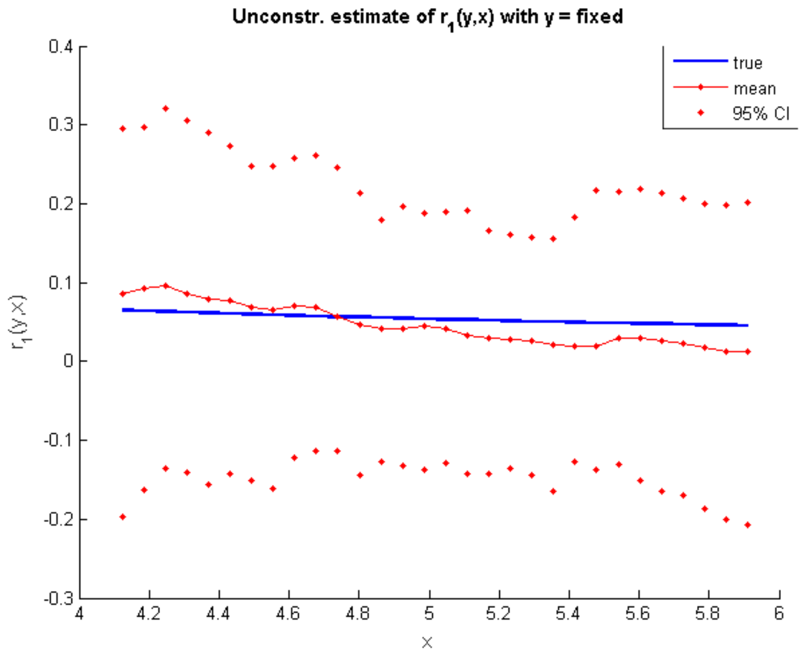

Figure 3: Estimation of $r_1(x, y)$ with $y = \bar{y}_1$ fixed.

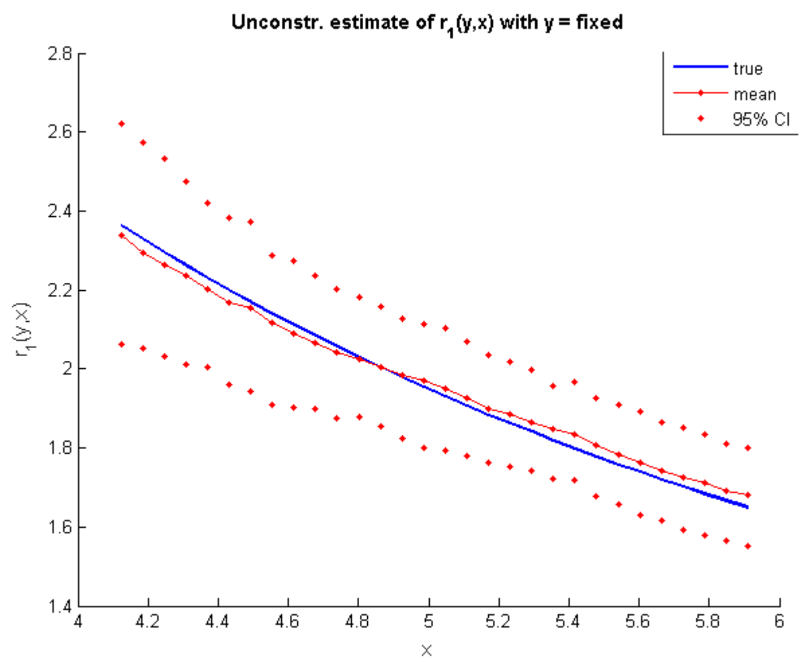

Figure 4: Estimation of $r_1(x, y)$ with $y = \bar{y}_2$ fixed.

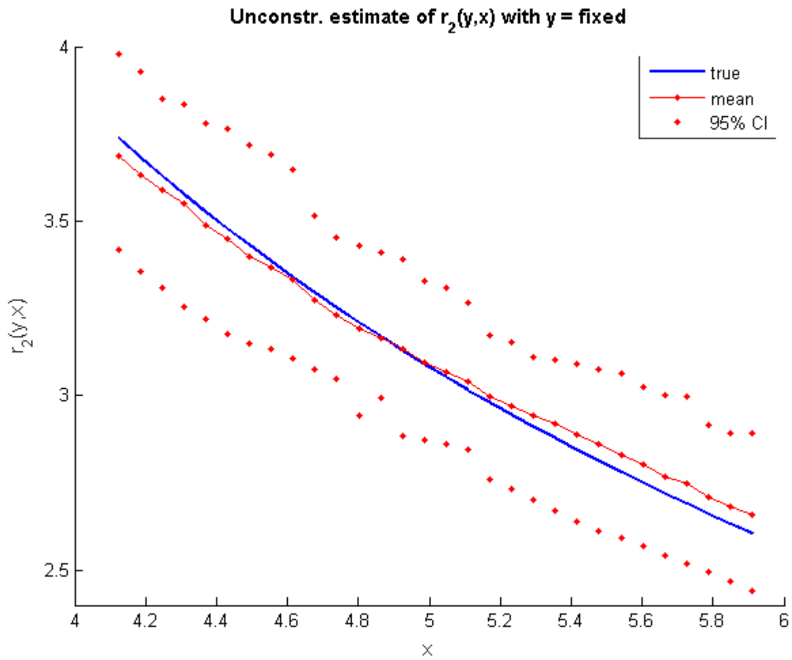

Figure 5: Estimation of $r_2\left(x,y\right)$ with $y=\bar{y}_1$ fixed.

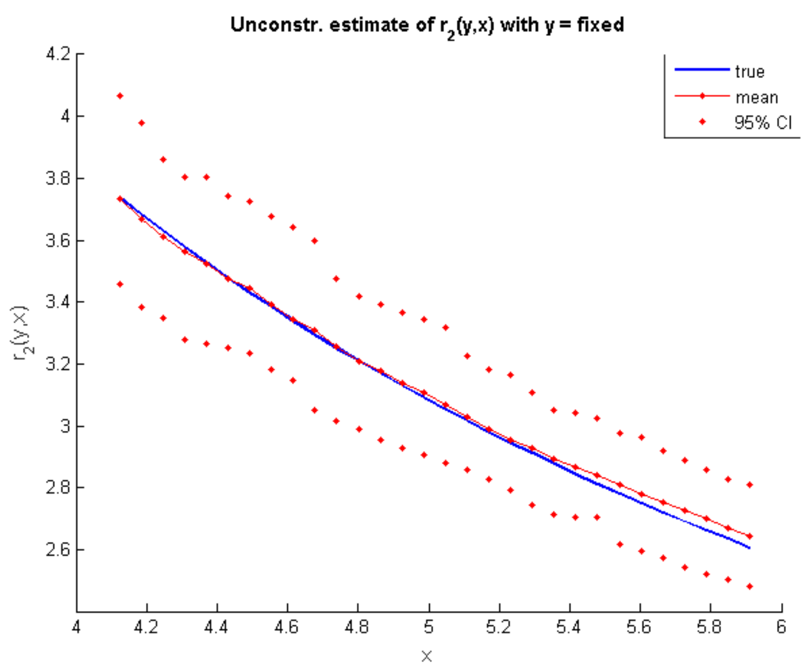

Figure 6: Estimation of $r_2\left(x,y\right)$ with $y=\bar{y}_2$ fixed.

D Example satisfying Assumption 3.2

For a particular example of a conditional density satisfying Assumption 3.2, suppose that $E\subset R_+^2$, and for a known 1-1 function λ from Z to R_{++}, the set of positive values in R,

$$f_{\varepsilon|Z}\left(\varepsilon_1,\varepsilon_2|z_i\right)=\left[\lambda(z_i)\right]^3\quad\exp\left(-\left[\lambda(z_i)\right]\ \varepsilon_1-\left[\lambda(z_i)\right]^2\ \varepsilon_2\right)$$

Then,

$$f_{\varepsilon|Z}\left(\varepsilon_1, \varepsilon_2|z_1\right) = f_{\varepsilon|Z}\left(\varepsilon_1, \varepsilon_2|z_2\right) = f_{\varepsilon|Z}\left(\varepsilon_1, \varepsilon_2|z_3\right)$$

if and only if

$$(3.1) \qquad 3\ \ln\left(\frac{\lambda(z_1)}{\lambda(z_2)}\right) - [\lambda(z_1) - \lambda(z_2)]\ \ \varepsilon_1 - [\lambda(z_1) - \lambda(z_2)]\,[\lambda(z_1) + \lambda(z_2)]\ \ \varepsilon_2 = 0$$

and

$$(3.2) \qquad 3\ \ln\left(\frac{\lambda(z_1)}{\lambda(z_3)}\right) - [\lambda(z_1) - \lambda(z_3)]\ \ \varepsilon_1 - [\lambda(z_1) - \lambda(z_3)]\,[\lambda(z_1) + \lambda(z_3)]\ \ \varepsilon_2 = 0$$

where we have used the equality $\left([\lambda(z_i)]^2 - [\lambda(z_j)]^2\right) = (\lambda(z_i) - \lambda(z_j))\,(\lambda(z_i) + \lambda(z_j))$. A unique solution $(\varepsilon_1, \varepsilon_2)$ exists to this system of two linear equations if and only if z_2 and z_3 are such that

$$\lambda(z_2) \neq \lambda(z_3)$$

Hence, since λ is a 1-1 function, it follows that as long as $z_2 \neq z_3$, there will exists a unique value $(\varepsilon_1, \varepsilon_2)$.