On Specification and Identification of Stochastic Demand Models

Walter Beckert*
Department of Economics
University of California, Berkeley

March 27, 2000

Abstract

This paper is concerned with stochastic demand systems that arise from structural random utility models for J continuous choice variables. It examines under which conditions on the structural preference specification the implied reduced form model induces choice vectors for $J-1$ inside goods whose distribution is non-degenerate, given prices and expenditure. And it investigates the conditions under which the structural model can be identified form demand data. Assessing the distribution of preferences is important when optimal demand management policies have to take into account heterogeneity in consumers' tastes. Model non-degeneracy is essential if the model design is not to impose a priori restrictions on behavior. The question of identification is fundamental in any structural approach. The paper shows that analogues to conventional assumptions on preferences, together with some easily verifiable further conditions, provide enough structure for identification through conditional moments, which can be derived from either the reduced form or, if analytically intractable, first-order conditions.

*I thank Chunrong Ai, Daniel McFadden, Lea Popovic, Paul Ruud and Chris Shannon for helpful comments and discussions. All errors are mine.

1 Introduction

Demand analysis is an actively researched area, in theoretical and, more prominently, applied work. The task of the applied researcher is to model randomness in demand data, given prices and income or expenditure. There exist, in principle, three paradigms to accomplish this task. The first attributes randomness in demand data to measurement error, the second interprets randomness as optimization error, and the third views randomness as arising from unobserved preference heterogeneity. These potential sources of randomness may contribute simultaneously. This paper focuses on the third interpretation. Identifying unobserved preference heterogeneity as the source of interest for randomness in demand data invokes a structural approach for econometric demand analysis. The elementary structural component in this approach is a random utility or stochastic preference model. This paper addresses two fundamental questions. Under which conditions on the specification of structural stochastic preferences over J goods does the implied reduced form induce choice vectors for the $J-1$ inside goods that have a non-degenerate distribution on the commodity space, given prices and expenditure? And under which conditions is the structural model identifiable from demand data? Model non-degeneracy is desirable in applied work if the modeling framework for analysis is not to impose a priori restrictions on consumers' behavior. The question of identifiability is critical in any structural, rather than reduced-form approach.

A structural random utility approach for micro-econometric demand analyses is of interest for several reasons. First, modeling unobserved preferences heterogeneity frees the empirical analysis from the rigidity of deterministic preferences. Typically, applied demand analyses maintain the hypothesis that consumers make rational choices that are consistent with a preference ordering on the commodity space. Empirically, however, even precisely measured choice data often do not satisfy rationality tests like the Weak Axiom of Revealed Preference.[1] The modeling assumption of deterministic preferences becomes a liability when there is evidence that consumers indeed optimize, because then the failure of the data to satisfy the Weak Axiom suggests that consumers' preferences are inconsistent across choice situations. A structural framework that allows for unobserved preference heterogeneity is flexible enough to accommo-

[1] The Weak Axiom is a weak consistency requirement on choices. Let $\mathbf{x}(\mathbf{p})$ and $\mathbf{x}(\tilde{\mathbf{p}})$ denote the choice vectors at prices $\mathbf{p}$ and $\tilde{\mathbf{p}}$, respectively. Then the Weak Axiom states that if $\mathbf{x}(\tilde{\mathbf{p}})$ was feasible when $\mathbf{x}(\mathbf{p})$ was chosen, i.e. if $\mathbf{x}(\tilde{\mathbf{p}})'\mathbf{p} \leq \mathbf{x}(\mathbf{p})'\mathbf{p}$, then it must not be that $\mathbf{x}(\mathbf{p})$ was feasible when $\mathbf{x}(\tilde{\mathbf{p}})$ was chosen, i.e. it must be the case that $\mathbf{x}(\tilde{\mathbf{p}})'\tilde{\mathbf{p}} < \mathbf{x}(\mathbf{p})'\tilde{\mathbf{p}}$.

date both, the maintained hypothesis of optimization in accordance with a consistent preference ordering and the empirically observed patterns in the data.

Second, next to such conceptual considerations, assessing the entire distribution of preferences is of practical importance when consumers compete for shared resources. Typical examples are the consumption of services that are delivered through capacity-constrained networks.[2] If aggregate consumption is close to exhausting capacity, then the service valuation and utilization of a marginal user potentially inflicts a negative consumption externality on all contemporaneous users, through network congestion and ensuing quality-of-service degradation. Calibrating consumers' taste heterogeneity is therefore instrumental for effective and efficient management of capacity-constrained resources.

Quality of service can be enhanced through efficient capacity allocation. At least in theory, efficient capacity allocation can be achieved through nonlinear pricing schemes for such services. The literature on the theory of nonlinear pricing offers preference heterogeneity as the fundamental motivation for the welfare-enhancing effects of nonlinearities in price schedules; see, for example, Wilson (1993). Heterogeneous preferences are a primitive in this theory. The third reason for measuring preference heterogeneity, then, is the desire to empirically assess the primitives of the theory of nonlinear pricing, in order implement its prescriptions for optimal price structures.

The literature on random utility models has primarily focused on discrete choice models. It goes back to Thurstone (1927) and McFadden (1974, 1981); see also McFadden and Train (1998) for some recent developments. The literature on stochastic preferences models for continuous choices is much more sparse. An early attempt to introduce randomness into the utility framework is Barten (1968), expanding a demand theory developed by Barten (1964) and Theil (1965, 1967). In applied work, Dubin and McFadden (1974) in an analysis of electricity demand and Fuss, McFadden and Mundlak (1978) in production analysis estimate structural econometric models with random parameters. More recently, Brown and Matzkin (1995, 1995A) consider a special class of non-parametric random utility models, give conditions for identification and propose a non-parametric estimation methodology for this class. Brown and Walker (1989) and Lewbel (1996) investigate the implications of the utility maximization hypothesis for the

[2] A classical instance of this kind is electricity. More topical examples are Internet access and wireless networks for voice and data.

reduced form system of stochastic demand functions if randomness in this system is interpreted as arising from unobserved preference heterogeneity.

The focus of the econometric analysis in this paper is the structural, parametric random utility model for continuous choices. This paper first provides a brief synopsis of the alternatives that are available to the applied researcher for modeling demand when randomness is interpreted as arising from unobserved preference heterogeneity. The paper then examines the necessary and sufficient conditions that a random preference specification must satisfy to generate a system of demands for $J-1$ inside goods that has a non-degenerate distribution on its support. Non-degeneracy is desirable if the model should be capable of statistically explaining any conceivable observation in the commodity space, given prices and expenditure. If no a priori restrictions on choice behavior are imposed by the model design, then the the distribution of demand for inside goods will be non-degenerate. If such restrictions were imposed and not valid, then estimates from a restricted model would be inconsistent. If an unrestricted model is estimated, then actual constraints on choice behavior can still be detected, as they manifest themselves in estimates of higher-order moments.

The answer to the question about non-degeneracy essentially focuses on the specification of the dimension of stochastic components in the random preference model that is necessary to model responses without restricting behavior through the choice of specification. It is shown that, under continuity and smoothness assumptions, the vector of observed response variables has a non-degenerate distribution on its support only if its dimension is not larger than the dimension of the stochastic preference component vector. A converse to this result gives easily verifiable conditions on random utility functions that are sufficient to guarantee that the distribution of demands is non-degenerate, if the necessary dimension requirement on the stochastic components is met. These conditions restrict random utility so that, as a result, 1) the Jacobian of the system of stochastic demand functions with respect to the stochastic components exists, and 2) the Jacobian has full column rank with probability one, given prices and income. If the conditions are slightly strengthened, then they can be shown to imply that the stochastic demand system is invertible almost surely. It is possible to have random preference components of dimensions higher than necessary, but, as illustrated in an example, identification in such situations is tenuous.

The second part of the paper turns to identification. It is shown that the conditions on the stochastic preference specification that suffice for an almost surely invertible model ensure that fixed parameters in the random preference model as well as the variance-covariance matrix of the random preference components are locally identifiable from demand data. Due to the nonlinearity typical in the functional relationship between response variables and stochastic preference components, the likelihood function contains an analytically intractable Jacobian term. As a consequence, estimation and identification strategies based on moment conditions are considered. It is, therefore, quite natural that the identification question asked in this context is intimately linked to the question of identification in generalized method of moments (GMM) setups. The proof of the main identification proposition, Propositions 1, appeals to this connection. It builds on conditional moments derived from the implied reduced form model. A corollary to the proposition reveals that moment conditions can be derived alternatively from the first-order conditions. This strengthens the identification result since it makes it applicable also in cases when the system of stochastic demand functions exists in theory only, but is analytically intractable in practice. The ensemble of the specification and identification results show that rather conventional assumptions about preferences that one might maintain in applied microeconomic consumption analyses provide enough structure for identification. Thus, identification really arises from the economics of the choice problem that generates the observed consumption data. Estimation of an identifiable model still requires appropriate moment restrictions. In fact, an example shows that even in a simple case with a single stochastic parameter and essentially two estimable parameters of interest, moments beyond the first two may be needed to estimate these parameters consistently.

The paper proceeds as follows. The second section derives necessary and sufficient conditions that the specification of stochastic preferences must satisfy to yield a non-degenerate distribution of demands for inside goods, conditional on prices and expenditure. It then turns to conditions under which the reduced form is alost surely invertible. It concludes with a discussion of random utility models with stochastic components of dimensions higher than necessary. The third section develops the results on identification of the fixed parameters of the preference specification and the variance-covariance matrix of the stochastic preference components from demand data. Various examples illustrate these results throughout the exposition. The final section summarizes and suggests some directions for future research.

2 Specification

2.1 Synopsis of different approaches

Suppose one observes consumption amounts for $J-1$ commodities, $\mathbf{x}_{-J}$, together with prices $\mathbf{p}$, including the price (index) of a (composite) outside good J, and total expenditure m. There are basically three different approaches to designing a stochastic demand model for the data at hand.

The first approach develops a stochastic reduced form, i.e. a system of stochastic demand equations for the observed consumption vector $\mathbf{x}_{-J}$, given prices $\mathbf{p} = (p_1, \ldots, p_J)'$ and income, assuming for identification purposes that realized demands emerge from demand functions, not correspondences. It is known from the work by Brown and Walker (1989) that the properties of the demand functions imposed by microeconomic consumer theory, namely homogeneity of degree zero, symmetry and negative semi-definiteness of the Slutsky matrix, conditional on prices and income, imply that the difference between realized demand and its conditional expectation is itself a function of prices and income. This implies that any consistent specification of stochastic demand equations must not have a sole, additive homoskedastic error term, but instead exhibit stochastic price and income derivatives. Any specification of the system of stochastic demand functions must in addition respect the natural boundaries for the dependent variables, amounts consumed or expenditure shares, by suitably limiting the support of the stochastic components. Specifications with homoskedastic additive error components are therefore only consistent with the interpretation of these error terms as arising from measurement error, not preference heterogeneity. Any reduced form specification has the drawback that it yields Marshallian, not Hicksian demand functions. This means that any marginal effects trace out Marshallian substitution effects, holding income constant, not income-compensated substitution along an indifference curve, so that it is not clear how to construct preferences from these demand equations. Moreover, it is not clear how to impose the restrictions implied by utility maximization. A variant of the first approach is a random coefficient specification of the demand functions. This approach is, however, equally unsatisfying since it is also unclear how to incorporate the restrictions implied by the utility maximization hypothesis.

One might argue – and it has been suggested by some, e.g. Lancaster (1966), p. 132, to whom the following "tongue in cheek" citation is due – that "determinateness of the sign of the substitution effect [is] the only substantive result of the theory of consumer behavior." Preferences are of interest not so much because they allow consumption bundles to be ranked, but more fundamentally because they characterize substitution relationships between goods. Once the substitution pattern is determined, at least in principle ordinal rankings can be inferred by integration. Another approach to stochastic demand analysis, therefore, might start from a stochastic specification of $J-1$ marginal rates of substitution, assuming continuously differentiable utility functions and quasi-concavity restrictions. This is in principle an attractive venue since marginal rates of substitution are observable from data on demands and relative prices. Compared to the approach of specifying the system of demand functions, it benefits from the fact that the stochastic specification of marginal rates of substitution must only respect their non-negativity. If structural preferences are identifiable form first-order conditions, then, however, there is no reason not to pursue a fully structural, random utility approach. As the third section of this paper shows, the conditions that yield identification from first-order conditions are the same as the ones that yield identification from the reduced form.

Therefore, one might as well consider the third, fully structural approach which specifies stochastic preferences directly by means of random utility functions defined over all J goods. Under suitable differentiability, monotonicity and quasi-concavity assumptions, this approach yields both, stochastic marginal rates of substitution and stochastic demand functions. While this is the cleanest modeling approach and allows the greatest flexibility in terms of the stochastic specification, it comes at the cost that parameters of interest in the random utility function must be identifiable from the implied $J-1$ stochastic demand functions for the observed consumption in the first $J-1$ goods, or, alternatively, from the first-order conditions. As will be shown below, this can be achieved through a specification which yields a stochastic demand system which is invertible for all values of prices and income.

2.2 Stochastic Dimensionality

This synopsis points to the fact that a general specification must be estimable from the implied system of demand functions or first-order conditions. Since this is so, any sensible stochastic model that is a candidate for estimation must be representable – at least in principle – as a system of demand functions that is consistent with the model and with general implications of microeconomic consumer theory and that does not a priori limit the space of observables. To formalize the latter requirement, consider the following

Definition: The random variable $\mathbf{x} \in \mathbb{R}^J$ has dimension J, denoted throughout by $\dim(\mathbf{x}) = J$, if it has a non-degenerate distribution on $\mathbb{R}^J$.[3]

Definition: Define a stochastic model for demand data on consumption, prices and income, $(\mathbf{x}'_{-J}, \mathbf{p}', m)' \in \mathbb{R}_+^{J-1} \times \mathbb{R}_{++}^J \times \mathbb{R}_+$, to be non-singular if it yields a system of stochastic Marshallian demand functions $h^M(\mathbf{p}, m; \theta, \epsilon) = (h(\mathbf{p}, m; \theta, \epsilon)', h_J(\mathbf{p}, m; \theta, \epsilon))'$ such that $\mathbf{x}_{-J} = h(\mathbf{p}, m; \theta, \epsilon)$ has dimension $J - 1$ for all $\mathbf{p}$ and m.

This definition requires that $\text{supp}(\mathbf{x}_{-J})$ have non-empty interior in $\mathbb{R}^{J-1}$, since otherwise $\mathbf{x}_{-J}$ lies on a manifold in $\mathbb{R}^{J-1}$ with dimension lower than $J - 1$ and does not have a non-degenerate distribution on $\mathbb{R}^{J-1}$. If the joint distribution of the vector $\mathbf{x}_{-J}$ has a singular variance-covariance matrix, then $\mathbf{x}_{-J}$ has dimension less than $J - 1$, but the converse is not necessarily true. The requirement of the definition is essential if, given prices and expenditure $(\mathbf{p}', m)' > \mathbf{0}$, the model should have the capability to explain every conceivable observation in the commodity space $\left[0, \frac{m}{p_1}\right] \times \ldots \times \left[0, \frac{m}{p_J}\right]$, subject to the linear budget constraint $\mathbf{p}'\mathbf{x} = m$. Restricting responses to a lower-dimensional manifold of $\mathbb{R}^{J-1}$ amounts to an a priori restriction of behavior by model design. An example below illustrates this point. For applied demand analyses, it seems preferable to first consider a structural model without such a restriction on behavior. If the restriction were in fact true, then it would manifest itself in estimates of higher-order moments. If it is wrong, yet imposed by the model, then model estimates are inconsistent; on the other hand, if the restriction is true, then imposing the restriction yields more efficient estimates. Therefore, as a modeling strategy, one would estimate an unrestricted model first, test for the validity of restrictions, and if, they are not rejected, one might proceed by estimate the restricted model.

Since, given $\mathbf{p}$ and m, the response variables $\mathbf{x}_{-J}$ are in a $J - 1$ dimensional subset of $\mathbb{R}_+^{J-1}$, it is at least plausible that the stochastic components ϵ of non-singular stochastic models should have $\dim(\epsilon) = J - 1$. Lemma 2 below states that this dimensionality requirement, in conjunction with conditions that guarantee the existence and smoothness of a system of demand functions and a Jacobian with respect to ϵ that has full column rank, is in fact a sufficient

[3] For simplicity, there is no notational distinction between the random variable and its realization.

condition for non-singular structural specifications of stochastic demand models. Under its additional assumptions, it can be viewed as a converse to the following general requirement on the stochastic dimensionality of the domain of Lipschitz and C^1 maps in order for such maps to induce a non-degenerate response distribution.

Lemma 1: Consider the map $x : \mathcal{E} \to \mathcal{X}$, for the vector-valued random variable ϵ with $E[\epsilon] = \mathbf{0}$, $E[\epsilon\epsilon'] = \Sigma$ and $\dim(\epsilon) = K$ and $x(\epsilon) \in \mathbb{R}^J$ for all $\epsilon \in \mathcal{E}$. If the map x is non-constant and either *(i)* uniformly Lipschitz continuous or *(ii)* C^1, then a necessary condition for $x(\epsilon)$ to have a non-degenerate distribution on $\mathbb{R}^J$ is that $K \geq J$.

The proof of Lemma 1 is given in an appendix. The lemma implies for structural random utility models with J continuous choice variables that their specification has to involve a stochastic component vector ϵ of at least stochastic dimension $J-1$ if the implied reduced form stochastic demand model is either continuously differentiable or Lipschitz continuous in ϵ. Continuous differentiability of the reduced form with respect to ϵ is convenient for a number of reasons. First, as Lemma 2 will show, it can be justified easily from assumptions on the structural preference specification. Second, under conditions on the structural model with $J-1$-dimensional ϵ that can be checked in a straightforward fashion in applications, the Jacobian of the reduced form system with respect to ϵ has full column rank $J-1$ almost surely. Together with ϵ having stochastic dimension $J-1$, this implies that the $J-1$-component choice vector of the $J-1$ inside goods, $\mathbf{x}_{-J}$, has a non-degenerate distribution. This is a necessary, though not sufficient condition for the model to be capable to explain all conceivably observable responses in the commodity space by a realization of the stochastic model component ϵ. Third, the existence of an almost surely non-singular Jacobian matrix allows to invoke the Inverse Function and Implicit Function theorems. Lemma 3 in the following subsection shows how these can be used to establish that the stochastic demand vector for the inside goods $\mathbf{x}_{-J}$ and the stochastic component vector ϵ are one-to-one almost surely. Thus, under suitable assumptions the model is invertible for any conceivable observation in the commodity space. Moreover, this property of the reduced form is used later to argue that structural parameters of random utility models that imply such reduced form models are identifiable on the basis of demand data.

Before stating Lemmas 2 and 3, it is necessary to introduce the following formalism. Denote by $\mathcal{U}$ the collection of all random (direct) utility functions $U : \mathbb{R}_+^J \times \Theta \times \mathbb{R}^K \to \mathbb{R}$, defined up to a finite dimensional parameter vector $\theta \in \Theta$ and a K dimensional random component

$\epsilon \in \mathbb{R}^K$, that are continuous in their arguments, continuously differentiable in ϵ and, for each ϵ, in $\mathbf{x} \in \mathbb{R}^J$ and θ, concave and strictly quasi-concave in $\mathbf{x}$ for each ϵ, such that no two functions are strictly increasing transformations of each other. Leaving aside differentiability, these assumptions are more than sufficient to guarantee that a system of Marshallian demand functions $h^M(\mathbf{p}, m; \theta, \epsilon) = (h(\mathbf{p}, m; \theta, \epsilon)', h_J(\mathbf{p}, m; \theta, \epsilon))'$ exists. For measurable representatives of equivalence classes[4] in the space of random utility functions, they essentially mimic the ones given in Mas-Colell (1977) for each ϵ and guarantee that, for each ϵ, a particular demand function cannot be generated by several different utility functions in $\mathcal{U}$. Differentiability is added so that marginal rates of substitution exist. The ensemble of these conditions shall henceforth be invoked whenever it is supposed that $U \in \mathcal{U}$. Assuming first and second moments of ϵ exist as well, without loss of generality let $E[\epsilon] = 0$, $E[\epsilon\epsilon'] = \Sigma$, $\mathrm{rk}(\Sigma) = K$.

Lemma 2 gives conditions on random utility specifications that are sufficient for model non-singularity in the sense of the definition above.

Lemma 2: *Consider $U \in \mathcal{U}$ where $U = U(\mathbf{x}, \theta, \epsilon)$, $\mathbf{x} \in \mathbb{R}^J_+$ and dim$(\epsilon) = J - 1$, $J \geq 2$. Suppose that (A1) the $J \times (J-1)$ matrix $\nabla_{\mathbf{x}\epsilon'} U(\mathbf{x}, \theta, \epsilon)$ has full column rank $J - 1$ for all θ, except on a set with probability zero, and that*

$$\begin{vmatrix} \nabla_{\mathbf{w}\mathbf{w}'} U(\mathbf{x}, \theta, \epsilon) & \nabla_{\mathbf{w}} U(\mathbf{x}, \theta, \epsilon) \\ \nabla_{\mathbf{w}'} U(\mathbf{x}, \theta, \epsilon) & 0 \end{vmatrix} \neq 0$$

for all θ and $\mathbf{w}' = (\mathbf{x}', \epsilon')$. Denote the system of Marshallian demand functions by

$$h^M(\mathbf{p}, m, \theta, \epsilon) = [h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon)]' = \arg\max_{\mathbf{x} \in \mathbb{R}^J_+} \{U(\mathbf{x}, \theta, \epsilon) : \mathbf{p}'\mathbf{x} = m\}.$$

Then, $\mathbf{x}_{-\mathbf{J}} = h(\mathbf{p}, m, \theta, \epsilon)$ has dim$(\mathbf{x}_{-J}) = J - 1$.

This result is illustrated by the two examples, stochastic parameter Cobb-Douglas preferences and linear expenditure systems. Stochastic parameter Cobb-Douglas preferences have recently been used in Beckert (1999). They are a special case of linear expenditure systems introduced by Klein and Rubin (1947). Early empirical applications are Stone (1954) and Pollak et al. (1969). Linear expenditure systems enjoy continued popularity in applied work; see, for example, Darrough et al. (1983) and Kapteyn et al. (1997).

[4] For conditions under which measurable utility functions exist, see Aumann (1969), Wesley (1976) and Wieczorek (1980).

Example: (*Stochastic Cobb-Douglas Preferences 1*)

Suppose that $U(\mathbf{x},\theta,\epsilon)=\sum_{j=1}^{J-1}e^{\theta_j+\epsilon_j}\ln(x_j)+\ln(x_J)$. The stochastic demand functions for the inside goods are

$$h(\mathbf{p},m,\theta,\epsilon)_k=\frac{e^{\theta_k+\epsilon_k}}{1+\sum_{j=1}^{J-1}e^{\theta_j+\epsilon_j}}\frac{m}{p_k},\text{ for k=1, \dots ,J-1,}$$

and the Jacobian is

$$\begin{aligned}\nabla_\epsilon h(\mathbf{p},m,\theta,\epsilon) &= \operatorname{diag}\left(h(\mathbf{p},m,\theta,\epsilon)_1,\dots,h(\mathbf{p},m,\theta,\epsilon)_{J-1}\right)\times \\ &\quad\left(\frac{-1}{1+\sum_{j=1}^{J-1}e^{\theta_j+\epsilon_j}}\operatorname{diag}\left(e^{\epsilon_1},\dots,e^{\epsilon_{J-1}}\right)\iota\iota'+\mathbf{I}_{J-1}\right),\end{aligned} \tag{2-1}$$

where $\iota=(1,\dots,1)'$, with $\iota'\iota=J-1$. The rank of this matrix is $J-1$, except possibly on sets of measure zero. This can alternatively be concluded from Lemma 1 if *A1* is verified. Differentiability is obvious, so the condition on the bordered Hessian need not be checked. To verify the rank assumption, in this case the matrix $\nabla_{\mathbf{x}\epsilon}U$ is given by

$$\nabla_{\mathbf{x}\epsilon}U(\mathbf{x},\theta,\epsilon)=\begin{bmatrix} e^{\theta_1+\epsilon_1}\frac{1}{x_1} & 0 & \cdots\cdots & \\ 0 & e^{\theta_2+\epsilon_2}\frac{1}{x_2} & 0 & \cdots \\ 0 & 0 & \ddots & \cdots \\ \vdots & & & \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

This matrix has dimension $J\times(J-1)$ and rank $J-1$ for all ϵ and $\mathbf{x}$. □

The following example shows that it is important to consider all equivalent representations of any specification of U in $\mathcal{U}$ to guard against model singularities. It also serves as an example of a model specification that can only explain observations that lie in a subspace of the commodity space and thus, effectively, imposes a restriction on behavior.

Example: (*Linear Expenditure System 1*)

Consider the stochastic preference model $U(\mathbf{x},\theta,\epsilon)=e^{\theta_1}\ln(x_1-\theta_2-e^{\epsilon_1})+e^{\theta_3+\epsilon_2}\ln(x_1-\theta_4)+e^{\epsilon_2}\ln(x_3-\theta_5)$, where $\theta\in\mathbb{R}^5_+$. For this model, the stochastic demand functions for the inside goods x_1 and x_2 are

$$\begin{aligned} h_1(\mathbf{p}, m, \theta, \epsilon) &= \theta_2 + e^{\epsilon_1} + \frac{e^{\theta_1}}{e^{\epsilon_2} + e^{\theta_1} + e^{\theta_3+\epsilon_2}} \left[\frac{m}{p_1} - \theta_4 \frac{p_2}{p_1} - \theta_5 \frac{p_3}{p_1} - \theta_2 - e^{\epsilon_1} \right] \\ h_2(\mathbf{p}, m, \theta, \epsilon) &= \theta_4 + \frac{e^{\theta_3+\epsilon_2}}{e^{\epsilon_2} + e^{\theta_1}} \left[\frac{m}{p_2} - (\theta_2 + e^{\epsilon_1}) \frac{p_1}{p_2} - \theta_5 \frac{p_3}{p_2} - \theta_4 \right]. \end{aligned}$$

Differentiability again is obvious. Checking the rank assumption (*A1*) first, one finds that

$$\nabla_{\mathbf{x}\epsilon} U(\mathbf{x}, \theta, \epsilon) = \begin{bmatrix} \frac{-e^{\theta_1+\epsilon_1}}{x_1-\theta_2-e^{\epsilon_1}} & 0 \\ 0 & \frac{e^{\theta_3+\epsilon_2}}{y-\theta_4} \\ 0 & \frac{e^{\epsilon_2}}{z-\theta_5} \end{bmatrix},$$

and so could be inclined to conclude - erroneously - that $\mathrm{rk}(\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)) = 2$ for the bivariate system $h(\mathbf{p}, m, \theta, \epsilon)) = (h_1(\mathbf{p}, m, \theta, \epsilon)), h_2(\mathbf{p}, m, \theta, \epsilon)))'$. It is easy to verify, however, that

$$\begin{aligned} \nabla_\epsilon h_1(\mathbf{p}, m, \theta, \epsilon) &= \begin{bmatrix} \frac{e^{\theta_1-\epsilon_2}(1+e^{\theta_3})}{(1+e^{\theta_1-\epsilon_2}+e^{\theta_3})^2} \left[\frac{m}{p_1} - \theta_4 \frac{p_2}{p_1} - \theta_5 \frac{p_3}{p_1} - \theta_2 - e^{\epsilon_1} \right] \\ \frac{e^{\epsilon_1}(1+e^{\theta_3})}{1+e^{\theta_1-\epsilon_2}+e^{\theta_3}} \end{bmatrix} \\ \nabla_\epsilon h_2(\mathbf{p}, m, \theta, \epsilon) &= \begin{bmatrix} -\frac{e^{\theta_3}(e^{\theta_1-\epsilon_2})}{(1+e^{\theta_1-\epsilon_2}+e^{\theta_3})^2} \left[\frac{m}{p_2} - (\theta_2 + e^{\epsilon_1}) \frac{p_1}{p_2} - \theta_5 \frac{p_3}{p_2} - \theta_4 \right] \\ -\frac{e^{\theta_3} e^{\epsilon_1}}{1+e^{\theta_1-\epsilon_2}+e^{\theta_3}} \frac{p_1}{p_2} \end{bmatrix}. \end{aligned}$$

Then, multiplying $\nabla_\epsilon h_1(\mathbf{p}, m, \theta, \epsilon)$ by $-\frac{e^{\theta_3}}{1+e^{\theta_3}} \frac{p_1}{p_2}$ produces $\nabla_\epsilon h_2(\mathbf{p}, m, \theta, \epsilon)$, so that the Jacobian has $\mathrm{rk}(\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)) = 1$ w.p.1.

This could have been detected immediately if one had considered the equivalent model $\tilde{U} \in \mathcal{U}$ with $\tilde{U}(\mathbf{x}, \theta, \epsilon) = e^{\theta_1-\epsilon_2} \ln(x_1 - \theta_2 - e^{\epsilon_1}) + e^{\theta_3} \ln(x_2 - \theta_4) + \ln(x_3 - \theta_5)$. For this model, the rank deficiency of the Jacobian follows immediately from Lemma 2, since

$$\nabla_{\mathbf{x}\epsilon} \tilde{U}(\mathbf{x}, \theta, \epsilon) = \begin{bmatrix} \frac{e^{\theta_1-\epsilon_2}}{x_1-\theta_2-e^{\epsilon_1}} & \frac{e^{\theta_1-\epsilon_2+\epsilon_1}}{x_1-\theta_2-e^{\epsilon_1}} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Form this, it can be seen that, conditional on prices, the model effectively restricts choices of x_2 and x_3 to the linear subspace $x_2 = \theta_4 + \frac{p_3}{p_2} e^{\theta_3} (x_3 - \theta_5)$. The model cannot explain observations that do not lie in this subspace. □

Within the class $\mathcal{U}$ of random utility functions with $K = J - 1$, one may wish to consider those for which x_j is non-separable from ϵ_j only, for $j = 1, \ldots, J - 1$. A special case of random utility functions of this form is given by $U(\mathbf{x}) + \mathbf{x}'\epsilon$, $\epsilon_J = 1$ w.p.1, a nonparametric model adopted in Brown and Matzkin (1995). The following result is a corollary to Lemma 1 and shows that this class of models yields systems of stochastic demand functions with non-degenerate distributions of the endogenous demand variables.

Corollary 1: *Suppose that $U(\mathbf{x}, \epsilon) \in \mathcal{U}$ satisfies $U(\mathbf{x}, \epsilon) = u(\mathbf{x}) + \mathbf{x}'\epsilon$, with $\epsilon_J = 1$ w.p.1 and dim$(\epsilon_{-J}) = J - 1$, and that $u(\mathbf{x})$ is smooth in the sense of Debreu.*[5] *Then, dim$(\mathbf{x}_{-J}) = J - 1$.*

Proof: Smoothness yields that the system of stochastic demand functions is continuously differentiable in it arguments, so that the Jacobian $\nabla_{\epsilon_{-J}} h^M(\mathbf{p}, m, \epsilon)$ exists. The vector of stochastic marginal utilities is

$$\nabla_{\mathbf{x}} U(\mathbf{x}, \epsilon) = \nabla_{\mathbf{x}} u(\mathbf{x}) + \epsilon,$$

so that $\nabla_{\mathbf{x}\epsilon'_{-J}} U(\mathbf{x}, \epsilon) = \mathbf{I}_{J-1}$. The result now follows by Lemma 1. □

2.3 Invertibility

Lemma 2 gives sufficient conditions for non-singularity of the reduced form model. These conditions are not enough, however, to ensure that, given $(\mathbf{p}', m)' > \mathbf{0}$ and θ, the image $\bigcup_{\epsilon \in \mathbb{R}^{J-1}} h(\mathbf{p}, m, \theta, \epsilon) = \text{Im}(h(\mathbf{p}, m, \theta, \cdot)) = \left(0, \frac{m}{p_1}\right) \times \ldots \times \left(0, \frac{m}{p_{J-1}}\right)$. In other words, the assumptions tied to the class $\mathcal{U}$ as well as assumption *A1* do not guarantee that any conceivable observation in the commodity space can be explained by a realization of ϵ. For this to be the case, the function $h(\mathbf{p}, m, \theta, \cdot)$, given $(\mathbf{p}', m)' > \mathbf{0}$ and θ, must be surjective. The following Lemma goes one step further. It gives conditions under which $h(\mathbf{p}, m, \theta, \cdot)$ is both surjective and injective so that the choice vector $\mathbf{x}_{-J}$ and ϵ stand in a one-to-one relationship.

Lemma 3: *Consider $U \in \mathcal{U}$, with $U = U(\mathbf{x}, \theta, \epsilon)$, $\mathbf{x} \in \mathbb{R}^J_+$ and dim$(\epsilon) = J - 1$. Suppose* A1 *holds, and that (*A1a*) $\text{Im}\left(\nabla_{\mathbf{x}} U(\mathbf{x}, \theta, \cdot)\right) = \bigcup_{\epsilon \in \mathbb{R}^{J-1}} \nabla_{\mathbf{x}} U(\mathbf{x}, \theta, \epsilon) = \mathbb{R}^J_{++}$, for all $\theta \in \Theta, \mathbf{x} \in \mathbb{R}^J_{++}$. Then, $\mathbf{x}_{-J} = h(\mathbf{p}, m, \theta, \epsilon)$ and ϵ are one-to-one, for all $(\mathbf{p}', m)' > \mathbf{0}$ and θ.*

The proof of Lemma 3 is in an appendix. The lemma implies that, for all $(\mathbf{p}', m)' > \mathbf{0}$ and $\theta \in \Theta$, the inverse function $g(\mathbf{p}, m, \theta, \cdot) := h^{-1}_{(\epsilon)}(\mathbf{p}, m, \theta, \cdot)$ exists, has $\mathbb{R}^{J-1}$ as its image and is single-valued for almost all $\mathbf{x}_{-J}$ in the commodity space.

[5] Debreu (1972); smoothness essentially amounts to the assumption on the determinant of the bordered Hessian, as in *A1*.

2.4 Higher-dimensional Specifications

Lemma 1 and 2 imply that a non-singular stochastic model for the envisaged demand data can adequately be built on at least $J-1$ stochastic components and satisfy the rank condition. It is also intuitively clear that models amenable to estimation, in principle, can involve ϵ with $\dim(\epsilon) > J-1$, provided that the rank condition is met so that the stochastic model specification is non-singular. Dimensions of the error vector higher than $J-1$, however, come at the cost that, in general, the density of the observable dependent variables $\mathbf{x}_{-J}$, conditional on $\mathbf{p}$ and m, cannot be computed from the density of ϵ. In this situation, the likelihood function of the parameters of interest, given the data, cannot be derived analytically. Moreover, if the model is identified, identification typically arises from a model specific nonlinearity. Since the model must be estimated from the conditional moments of $\mathbf{x}_{-J}$, this may be of limited use in practical estimation. The following example illustrates these points.

Example: (*Linear Expenditure System 2*)

Consider a random utility model of the form $U(x, y; \theta, \epsilon_x, \epsilon_y) = e^{\epsilon_x} \ln(x-\theta) + \ln(y - e^{\epsilon_y})$, for $\theta > 0$. Suppose that $(\epsilon_x, \epsilon_y)'$ are jointly normally distributed with mean 0 and nonsingular covariance matrix Σ. Observed data are $(x, p', m)'$. Then, the stochastic demand function for x is $x(p, m; \theta, \epsilon) = \theta + \frac{e^{\epsilon_x}}{1+e^{\epsilon_x}}\left[\frac{m}{p_x} - \theta - \frac{p_y}{p_x}e^{\epsilon_y}\right]$. The density of x, conditional on $(p', m)'$ and ϵ_y, is

$$\frac{|p_x(m-\theta-p_y e^{\epsilon_y} - p_x(x-\theta)|dx}{\sqrt{\Sigma_{11} - \Sigma_{12}^2/\Sigma_{22}}}\phi\left(\frac{\ln(p_x(x-\theta)) - \ln(m-\theta-p_z e^{\epsilon_y} - p_x(x-\theta)) - \frac{\Sigma_{12}}{\Sigma_{22}}\epsilon_y}{\sqrt{\Sigma_{11} - \Sigma_{12}^2/\Sigma_{22}}}\right),$$

where $\phi(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$, $u \in \mathbb{R}$. The parameters of this model are identified, because of the nonlinearity of the conditional expectation of ϵ_x, given $(p', m, \epsilon_y)'$, in ϵ_y. Since *epsilon*$_y$ cannot be integrated out, the model parameters must be estimated from the conditional moments of x by method-of-moments techniques[6]. The first conditional moment is

$$\begin{aligned} E[x|p,m] &= \frac{\theta}{2} + \frac{1}{2}\frac{m}{p_x} - \frac{p_y}{p_x}E\left[\frac{e^{\epsilon_x+\epsilon_y}}{1+e^{\epsilon}_x}\right] \\ &= \frac{\theta}{2} + \frac{1}{2}\frac{m}{p_x} - \frac{p_y}{p_x}e^{\frac{1}{2}(\Sigma_{22}-\Sigma_{12}^2/\Sigma_{11})}E\left[\frac{e^{\epsilon_x(1+\Sigma_{12}/\Sigma_{11})}}{1+e^{\epsilon_x}}\right] \\ &= \frac{\theta}{2} + \frac{1}{2}\frac{m}{p_x} - \frac{p_y}{p_x}\mu(\Sigma). \end{aligned}$$

[6] An alternative to MOM estimation would be to maximize the simulated likelihood function.

Thus, from the conditional mean of x, θ and $\mu(\Sigma)$ can be estimated by regressing $x - \frac{1}{2}\frac{m}{p_x}$ on a constant and $\frac{p_y}{p_x}$. To get estimates of the covariance parameters, it is necessary to resort of the conditional variance of $x - E[x|p,m]$ and to regress estimated errors from the first-stage regression on $\left(\frac{m^2}{p_x^2}, \frac{p_y^2}{p_x^2}, \frac{mp_y}{p_x^2}\right)'$. This follows from

$$\begin{aligned} E\left[(x - E[x|p,m])^2 \middle| p,m\right] &= \frac{m^2}{p_x^2}\left(E\left[\frac{e^{2\epsilon_x}}{(1+e^{\epsilon_x})^2}\right] - \frac{1}{4}\right) \\ &\quad + \frac{p_y^2}{p_x^2} e^{\Sigma_{22}-\Sigma_{12}^2/\Sigma_{11}} E\left[\frac{e^{2\epsilon_x(1+\Sigma_{12}/\Sigma_{11})}}{(1+e^{\epsilon_x})^2}\right] \\ &\quad - 2\frac{mp_y}{p_x^2}\left(e^{\frac{1}{2}(\Sigma_{22}-\Sigma_{12}^2/\Sigma_{11})} E\left[\frac{e^{2\epsilon_x(1+\Sigma_{12}/\Sigma_{11})}}{(1+e^{\epsilon_x})^2}\right] - \frac{1}{2}\mu(\Sigma)\right) \\ &= A(\Sigma_{11})\frac{m^2}{p_x^2} + B(\Sigma)\frac{p_y^2}{p_x^2} + C(\Sigma)\frac{mp_y}{p_x^2}. \end{aligned}$$

Given the estimates of θ and μ from the conditional mean regression, Σ_{11} can be identified from A, while $2B/(C + \frac{1}{2}\mu) = \frac{1}{2}(\Sigma_{22} - \Sigma_{12}^2/\Sigma_{11})$, so that Σ_{22} and Σ_{12} can now be identified from B, C and μ. Notice, however, if m and the price index of the composite outside good, p_y, grow (quite plausibly) approximately proportionately, then the design matrix of the second regression is ill-conditioned due to the approximate collinearity of its columns. In this case, although the model is theoretically identified, no stable estimates of the covariance parameters can be obtained from method-of-moments estimation. □

Having given conditions on the model dimensionality that are necessary and sufficient for model non-singularity, the question of identification of model parameters from demand data can be addressed. Identification is critical in any structural econometric approach. The following section shows that the structure imposed by consumer theory, augmented by assumptions *A1* and *A1a*, is rich enough for identification.

3 Identification

Consider random utility models in the class $\mathcal{U}$. In light of standard assumption about preference in economic theory, the stochastic preference components are assumed throughout to be independent of prices and income. The random preference components ϵ, in addition to satisfying the moment conditions $E[\epsilon] = \mathbf{0}$ and $E[\epsilon\epsilon'] = \Sigma$, is assumed to have stochastic dimension $\dim(\epsilon) = J - 1$. As before, denote the associated system of stochastic Marshallian demand functions by $\mathbf{x} = h^M(\mathbf{p}, m, \theta, \epsilon) = (h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))'$, given prices $\mathbf{p} \in \mathbb{R}^J_{++}$ and income $m > 0$, i.e. $h^M(\mathbf{p}, m, \theta, \epsilon) = \arg\max_{\mathbf{x}\in\mathbb{R}^J_+}\{U(\mathbf{x}, \theta, \epsilon) : \mathbf{x}'\mathbf{p} = m\}$; recall that strict monotonicity of $U \in \mathcal{U}$ for all θ and ϵ implies that consumption expenditure $\mathbf{p}'h^M(\mathbf{p}, m, \theta, \epsilon)$ equals income m.

Suppose that $U \in \mathcal{U}$ and assumption *A1* and *A1a* are satisfied. Then Lemma 1 implies for the Jacobian of the system of stochastic demand functions for the inside goods that $\text{rk}(\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)) = J-1$ for all $\mathbf{p}, m, \epsilon$. Thus, the vector or responses $\mathbf{x}_{-J}$ has a non-singular distribution of full dimension $J-1$. Assuming – for expositional purposes only – that ϵ has a density, the density of $\mathbf{x}_{-J}$, given $\mathbf{p}, m$ and parameterized by θ and $\mathbf{A}$, is given by

$$f_{\mathbf{x}_{-J}}(\mathbf{x}_{-J}|\mathbf{p}, m, \theta, \Sigma)d\mathbf{x}_{-J} = f_\epsilon\left(\Sigma^{-\frac{1}{2}}g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})\right)\left|\Sigma^{-\frac{1}{2}}\right|\left|\nabla_{\mathbf{x}_{-J}}g\left(\mathbf{p}, m, \theta, \mathbf{x}_{-J}\right)\right|d\mathbf{x}_{-J},$$

where $g(\mathbf{p}, m, \theta, \cdot) := h^{-1}_{(\epsilon)}(\mathbf{p}, m, \theta, \cdot)$ and where the Jacobian term arises from the changing variables due to the nonlinearity of the system $h^M(\mathbf{p}, m, \theta, \epsilon)$ in ϵ. From the definition of g, the Jacobian is given by $|\nabla_{\mathbf{x}_{-J}}g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})| = |\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)|^{-1}$. Notice that the Jacobian term $|\nabla_{\mathbf{x}_{-J}}g|$ depends on $\mathbf{x}_{-J}$ and $|\nabla_\epsilon h|$ on ϵ, respectively. This must be the case, since the assumption of $E[\epsilon] = \mathbf{0}$ does not allow $h(\mathbf{p}, m, \theta, \epsilon)$ to be linear in ϵ because in this case $P(h(\mathbf{p}, m, \theta, \epsilon) < 0) > 0$ since the support of ϵ is $\mathbb{R}^{J-1}$, while the support of $\mathbf{x}$ is the hyper-rectangle $\mathcal{X} = [0, \frac{m}{p_1}] \times \ldots \times [0, \frac{m}{p_J}]$. [7]

Denote the true parameters by θ_0 and Σ_0, respectively. Suppose that θ_0 and Σ_0 are not identified, i.e. that there exists $(\tilde{\theta}', \text{vec}(\tilde{\Sigma})')' \neq (\theta_0', \text{vec}(\Sigma_0)')'$, such that for all $\mathbf{p}, m, \mathbf{x}$,

$$f_{\mathbf{x}_{-J}}(\mathbf{x}_{-J}|\mathbf{p}, m, \theta_0, \Sigma_0)d\mathbf{x}_{-J} \quad = \quad f_{\mathbf{x}_{-J}}(\mathbf{x}_J|\mathbf{p}, m, \tilde{\theta}, \tilde{\Sigma})d\mathbf{x}_{-J}, \tag{3-2}$$

which is equivalent to

$$\begin{aligned} & f_\epsilon(\Sigma_0^{-\frac{1}{2}}g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J}))|\Sigma_0^{-\frac{1}{2}}||\nabla_{\mathbf{x}_{-J}}g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})|d\epsilon \\ = \quad & f_\epsilon(\tilde{\Sigma}^{-\frac{1}{2}}g(\mathbf{p}, m, \tilde{\theta}, \mathbf{x}_{-J}))|\tilde{\Sigma}^{-\frac{1}{2}}||\nabla_{\mathbf{x}_{-J}}g(\mathbf{p}, m, \tilde{\theta}, \mathbf{x}_{-J})|d\epsilon. \end{aligned} \tag{3-3}$$

Lack of identification means that there exists two distinct parameterizations that translate the distribution F_ϵ of the innovations ϵ into the same distribution of $\mathbf{x}_{-J}$, given $\mathbf{p}$ and m.[8]

[7] This, in part, distinguishes this work from Newey, Powell and Vella (1999) who consider triangular simultaneous equations models with additive, conditional mean-zero errors.

[8] Compare the case $x_t = \sigma\epsilon_t + \theta\sigma\epsilon_{t-1}$, for innovations ϵ_t being distributed independently with $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = 1$ for all t. This parameterization is non-identifiable since it translates the distribution of the innovations into the same distribution of x_t as the parameterization $x_t = \theta\sigma\epsilon_t + \sigma\epsilon_{t-1}$.

The following proposition provides sufficient conditions under which model parameters θ_0 and Σ_0 in random-parameter utility models are identified locally from demand data on consumption choices, prices and income $(\mathbf{x}'_{-J}, \mathbf{p}', m)'$. The proposition illustrates that microeconomic consumer theory, when augmented by distributional assumptions sufficient for non-degeneracy of the distribution of response variables, provides enough structure for identification. Its proof is outlined in an appendix.

Proposition 1: (*Random Parameter Utility Functions*)

Suppose that

A2 : the class of models $U \in \mathcal{U}$ *being considered satisfies* $U(\mathbf{x}, \theta, \epsilon) = U(\mathbf{x}, \theta_1 + \epsilon, \theta_2)$, *where* $\mathbf{x} \in \mathcal{X} := [0, \frac{m}{p_1}] \times \ldots \times [0, \frac{m}{p_J}]$, $\mathbf{p} \in \mathbb{R}^J_{++}$, $m > 0$, $\theta = (\theta_1', \theta_2')'$, $\theta_1 \in \Theta_1 \subset \mathbb{R}^{J-1}$,$\theta_2 \in \Theta_2 \subset \mathbb{R}^{K-(J-1)}$, *where* Θ_1 *and* Θ_2 *are open sets and* $J - 1 \leq K \leq 2(J-1)$;
$\nabla_{\mathbf{x}\theta_{1j}} U(\mathbf{x}, \theta, \epsilon) \not\propto \nabla_{\mathbf{x}\theta_{2k}} U(\mathbf{x}, \theta, \epsilon)$ *for all* $\mathbf{x} \in \mathcal{X}$, *on a set of positive probability,* $j = 1, \ldots, J-1; k = 1, \ldots, K-(J-1)$;
for $\epsilon \in \mathbb{R}^{J-1}$ *having dim*$(\epsilon) = J-1$ *and the moments* $E[\epsilon] = \mathbf{0}$, $E[\epsilon\epsilon'] = \Sigma_0$ *positive-definite, and* ϵ *and* $(\mathbf{p}', m)'$ *independent; let* $\theta_0 \in \Theta = \Theta_1 \times \Theta_2$ *denote the true parameter;*

A1' assumptions A1 and A1a hold, for $\mathbf{w} = (\mathbf{x}', \theta')'$;

A3: the distribution $F_{\mathbf{p}_{-J},m}$ *of* $(\mathbf{p}'_{-J}, m)'$ *is non-degenerate on* $\mathbb{R}^J_{++}$, *independent of* ϵ.

Then, the parameters θ_0 *and* Σ_0 *are locally identified*[9] *from data* $(\mathbf{x}'_{-J}, \mathbf{p}', m)'$.

Remarks: An alternative parameterization would have $U(\mathbf{x}, \theta_2, \epsilon)$, i.e. U parameterized by θ_2 and ϵ, with $E[\epsilon] = \theta_1$. The parameterization adopted in assumption *A2* of the proposition was chosen because it is more parsimonious in terms of notation. A second remark concerns the distributional assumptions. Microeconomic consumer theory axiomatically postulates that preferences do not depend on prices and income. This motivates the assumption of the stochastic preference components ϵ being independent of $\mathbf{p}_{-J}$ and m. Since only relative prices matter for consumption choices, p_J can be thought of as a normalizing constant.

[9]For a definition of local identification, see Rothenberg (1971). For a random variable U with parametric p.f. $f_U(u;\theta), \theta \in \Theta \subset \mathbb{R}^K, \theta$ is locally identified if for any other parameter vector $\tilde{\theta} \neq \theta$ in a neighborhood of θ the probability $\Pr(f_U(u;\theta) \neq f_U(u;\tilde{\theta}))$ is strictly positive.

Proposition 1 has the following intuitive interpretation. Suppose one wanted to estimate θ. Proposition 1 suggests that this could be done using a GMM estimator.[10] Moreover, the argument behind Proposition 1 suggests the interpretation of the GMM estimator as a *nonlinear instrumental-variable estimator*. Nonlinearity arises to the extent that $\nabla_\theta g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})$ may still depend on θ, unlike in the linear case. The result that $E[Z(\mathbf{p}, m, \theta_0)\nabla_\theta g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})]$ has full rank under assumptions *A1-3* can then be seen as analogous to the usual condition in the linear case that the expectation of the product of instruments and covariates be nonsingular. Furthermore, the distributional assumption *A3* implies that the matrix of instruments $Z(\mathbf{p}, m, \theta)$ has a nonsingular variance-covariance matrix. This is the analogue to the usual assumption in linear IV estimation that the second moment matrix of the array of instruments is nonsingular. Note that the instruments are functions of $\nabla_\theta h(\mathbf{p}, m, \theta_0, \epsilon)$; in fact, $Z(\mathbf{p}, m, \theta_0) = E[\nabla_\theta g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})|\mathbf{p}, m] = E[[\nabla_{\theta_1} h(\mathbf{p}, m, \theta, \epsilon)]^{-1}\nabla_\theta h(\mathbf{p}, m, \theta, \epsilon)|\mathbf{p}, m]$. As such, the instruments used in the identification argument minimize the asymptotic variance-covariance matrix of the nonlinear IV estimator since they minimize the mean squared error of the linear prediction of $\nabla_\theta g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})$ on the basis of predictors Z which satisfy $E[Z\nabla_\theta g]$ nonsingular and $E[Z'Z]$ nonsingular; compare, for example, Ruud (1999), Lemma 20.4 and Example 21.3. Thus, they constitute an array of efficient instruments.

The proof of Proposition 1 establishes the sufficient condition for identification, $(A-9)$, from the structure of preferences and the assumption that the vector $(\mathbf{p}'_{-J}, m)'$ has a non-degenerate distribution on $\mathbb{R}^{J-1}_{++}$. The following features of the structure of preferences are instrumental in proving the identification result. Essentially, the argument requires, on the one hand, that all the assumptions that yield smooth demand functions hold almost surely. Lack of strict quasi-concavity, for instance, eliminates smoothness; lack of strict monotonicity leads to demand correspondences. Beyond that, one needs in addition that, on a set of positive probability, any two utility parameters affect utility at the margin in different ways. The intuition for this requirement stems from the fact that parameters must be identifiable from the system of demands. This system is derived by means of the Implicit Function Theorem under the condition on the bordered Hessian (*A1'*), in which the utility gradient appears explicitly. If any two parameters affect the utility gradient in the same way almost surely, then this implies that,

[10]In fact, due to the nonlinearity of the function h in ϵ and of the inverse function g in $\mathbf{x}_{-J}$, the procedure would employ simulation to replace analytically intractable theoretical expectations by their simulated counterparts; in applications, the appropriate estimator therefore is a Method-of-Simulated-Moments estimator; compare, for example, Beckert (1999). The issue of simulation does not impede the interpretation given here.

w.p.1, these parameters also affect demands in the same way, given any $\mathbf{p}$ and m, and therefore are indistinguishable. In summary, Proposition 1 shows that identification really arises from the economics of the consumer choice problem that generates the observed responses $\mathbf{x}_{-J}$.

Example: (*Stochastic Cobb-Douglas Preferences 2*)

Suppose again that $U(\mathbf{x}, \theta, \epsilon) = \sum_{j=1}^{J-1} e^{\theta_j + \epsilon_j} \ln(x_j) + \ln(x_J)$, where ϵ satisfies *A2*. In this case, $\dim(\theta) = \dim(\epsilon) = J - 1$, so θ_2 is redundant. The stochastic demand functions for the inside goods are

$$h(\mathbf{p}, m, \theta, \epsilon)_k = \frac{e^{\theta_k + \epsilon_k}}{1 + \sum_{j=1}^{J-1} e^{\theta_j + \epsilon_j}} \frac{m}{p_k}, \text{ for k=1, \dots ,J-1,}$$

and the Jacobian is

$$\begin{aligned} \nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon) &= \operatorname{diag}\left(h(\mathbf{p}, m, \theta, \epsilon)_1, \dots, h(\mathbf{p}, m, \theta, \epsilon)_{J-1}\right) \times \\ &\quad \left(\frac{-1}{1 + \sum_{j=1}^{J-1} e^{\theta_j + \epsilon_j}} \operatorname{diag}\left(e^{\epsilon_1}, \dots, e^{\epsilon_{J-1}}\right) \iota\iota' + \mathbf{I}_{J-1} \right), \end{aligned} \tag{3-4}$$

where $\iota = (1, \dots, 1)'$, with $\iota'\iota = J - 1$. The rank of this matrix is $J - 1$, except possibly on sets of measure zero. This also follows from *A1*, as shown in the related example in the previous section. Thus, the model is invertible. Due to the random parameter model structure with $K = J - 1$, $\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon) = \nabla_\epsilon h(\mathbf{p}, m, \theta + \epsilon) = \nabla_\theta h(\mathbf{p}, m, \theta + \epsilon)$, and therefore $E[(\nabla_\epsilon h(\mathbf{p}, m, \theta + \epsilon))^{-1} \nabla_\theta h(\mathbf{p}, m, \theta + \epsilon)] = \mathbf{I}_{J-1}$, for all p, m and any θ. Thus,

$$E[g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] - E[g(\mathbf{p}, m, \tilde{\theta}, \mathbf{x}_{-J}] \approx -\mathbf{I}_{J-1}(\tilde{\theta} - \theta) = \theta_0 - \tilde{\theta},$$

and so θ_0 is identified. As already seen, Σ_0 can be identified from ($A - 8$). □

The identification argument for θ benefits from being closely related to the (simulation-assisted) GMM estimation methodology that one would use to uncover the structural model parameters from choice data. Conversely, the identification argument can guide estimation and enhance its efficiency since it points to efficient instruments. This distinguishes this line of reasoning from the one offered by Brown and Matzkin (1998)[11] whose identification and estimation approaches appear to be mutually uninformative.

The proof of Proposition 1 is stated in terms of the system of inverse demand functions $g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})$; compare ($A - 7$). It could have been equally derived in terms of the system

[11] Brown and Matzkin (1998), Theorem 3 and Corollary 1.

of demand functions $h(\mathbf{p}, m, \theta, \epsilon)$. One might therefore be led to believe that, for the purpose of GMM estimation of θ, it is necessary to be able to explicitly derive the system of stochastic demamd functions, as in the example above. This would limit the applicability of Proposition 1, since in many interesting cases $U \in \mathcal{U}$, this may not be possible. Even though in theory the existence of this system is guaranteed, there does not exist a close form representation for it. All that is analytically tractable are the usual first-order conditions. A variant of this approach, focusing primarily on the substitution pattern between commodities, has been pursued in the literature on intertemporal consumption decisions and essentially works with an intertemporal Euler condition, which amounts to the equality of an expected marginal rate of substitution and the relative price of consumption between periods, where both analyst and decision maker face the same uncertainty.[12] It follows, however, as a corollary to the proposition that the system of first-order conditions is absolutely sufficient for local identification of θ.

Corollary 2: *Suppose assumptions A1', A2 and A3 hold. Then θ is locally identifiable from the first-order conditions.*

The structure of the proof mimics the one of Proposition 1 and is given in an appendix.

This section concludes with some comments on estimation which are closely related to the insights developed up to this point. The above arguments underlying propositions 1 and 2 implicitly point to the fact that model parameters can be identified and therefore estimated from conditional and unconditional moments. Alternatively, they can in principle also be estimated by maximizing the log-likelihood function. Estimation from moment conditions may be more practical in applications, since this approach does not require the assumption of the stochastic utility components having a density. Also, compared to an approach based on the likelihood, moments of the response variables can be obtained directly, without recourse to the inverted model. The inverted model, however, is necessary to evaluate the density of the stochastic utility components. On the other hand, the likelihood-based approach yields an important by-product. Recall that Lemma 1 implies that model invertibility and a non-degenerate distribution of the $J-1$ vector ϵ are sufficient for a non-degenerate distribution of the vector of observed response variables $\mathbf{x}_{-J}$. In comparison to nonparametric models, parametric models therefore require that the Jacobian $\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)$ have full rank for all θ

[12]Compare, for example, Attanasio and Weber (1993). The stochastic component is often interpreted as arising from permanent and transitory income shocks, rather than from taste uncertainty.

in the parameter space, with probability one.[13] The maximum likelihood estimator imposes this requirement automatically, since, in light of (3 − 3), the logarithm of the Jacobian term approaches negative infinity for values of θ that render the Jacobian singular. This property of the ML estimator in this context is analogous to the fact that the ML estimator in autoregressive models, when based on the entire data set, restricts the coefficients of the innovation process to stationary.[14] It is an open question at this point how to impose non-singularity of the Jacobian for θ in method-of-moments estimation.

Care must be taken when implementing method-of-moments estimation. Due to model nonlinearity, there exist cases in which it does not suffice to have the same number of moment conditions and unknown parameters. The final example illustrates this in a model with two parameters that enter the estimating equations nonlinearly and in which the first two conditional moments do not suffice.

Example: (*Linear Expenditure System 3*)

Consider the variation of the previous model given by $U(x, y; \theta, \epsilon) = e^{\alpha+\sigma^2\epsilon} \ln(x - b_x) + \ln(y + b_y)$, where $\theta = (\alpha, \sigma^2, b_x, b_y)'$ are parameters of interest, which are positive except for α, and ϵ has a standard normal distribution. Observations are vectors $(x, \mathbf{p}', m)'$. Then, it follows from an argument analogous to the one given in the second example that the parameters of this model are identified. The conditional mean of x, given $\mathbf{p}, m$, is $E[x|\mathbf{p}, m] = b_x + \mu(\alpha, \sigma^2)(\frac{m}{p_x} - b_x - \frac{p_y}{p_x} b_y)$, where $\mu(\alpha, \sigma^2) = E\left[\frac{e^{\alpha+\sigma^2\epsilon}}{1+e^{\alpha+\sigma^2\epsilon}}|\mathbf{p}, m\right] = E\left[\frac{e^{\alpha+\sigma^2\epsilon}}{1+e^{\alpha+\sigma^2\epsilon}}\right]$ under the assumption of ϵ being independent of $\mathbf{p}$ and m. So identification of b_x and b_y from the conditional mean does not pose any problems. Identification of α and σ^2 from these moments alone, however, is not possible. To see this, consider the following properties of μ which are easily established for all α, σ^2: $\mu(\alpha, \sigma^2) = 1 - \mu(-\alpha, \sigma^2)$ $\frac{d}{d\alpha}\mu(\alpha, \sigma^2) > 0$, $\frac{d}{d\sigma^2}\mu(\alpha, \sigma^2) = 0$ if $\alpha = 0$, > 0 if $\alpha \ll 0$, < 0 if $\alpha \gg 0$. The last feature reflects the fact that, given a value of μ, the implied relationship between α and σ^2 becomes a correspondence as α approaches zero. The conditional variance of x is given by $\mathrm{var}(x|\mathbf{p}, m) = \left(\frac{m}{p_x - b_x} - \frac{p_y}{p_x} b_y\right)^2 \Sigma(\alpha, \sigma^2)$, where $\Sigma(\alpha, \sigma^2) = E\left[\left(\frac{e^{\alpha+\sigma^2\epsilon}}{1+e^{\alpha+\sigma^2\epsilon}} - \mu(\alpha, \sigma^2)\right)^2\right]$. It can be shown that, given a value of Σ, the implied relationship between α and σ^2 is approximately linear. This, however, means that the parameters of the model cannot be identified from the

[13] The Implicit Function Theorem then implies that $\nabla_{\mathbf{x}_{-J}} g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})$ has full rank for all θ and almost all $\mathbf{x}_{-J}$ in the opportunity set.

[14] Cp. Ruud (1999), section 25.2.2

first and second conditional moments alone. The nonlinearity in which ϵ appears in the demand equation is the source of this result. If the model is altered to $U(x, y; \theta, \epsilon) = e^{\alpha} \ln(x - b_x e^{\sigma^2 \epsilon}) + \ln(y + b_y)$, then the demand equation is linear in the error component, and first and second conditional moments suffice to identify the parameters of this model.

4 Conclusions

This paper presents framework for specifications of structural stochastic demand models in which randomness in demand data arises from preference heterogeneity. It provides necessary and sufficient conditions on the specification of random preference under which the vector of demands, conditional on prices and income, has a non-degenerate distribution and which are straightforward to check in applications. Furthermore, an identification theorem for natural specifications of parametric random utility models is derived, based on assumptions about the first two moments of the stochastic model components. The theorem and a corollary demonstrate that identification and estimation of structural model parameters is possible from generalized moment conditions that can be derived from the implied stochastic demand system or, alternatively, from the first-order conditions. The latter option renders the identification result more robust, as it makes it applicable to structural preference models whose implied stochastic demand system exists in theory, yet is analytically intractable. The identification results show that preferences themselves provide enough structure for identification. The economics of the consumer choice problem that generates the observed consumption data, thus, is seen to provide sufficient structure for identification. Several examples illustrate these results.

Future work might attempt to unify the framework adopted here – which is parametric in terms of the specification of random preference models, but only maintains moment assumptions on their stochastic model components – with the non-parametric work of Brown and Matzkin – which is non-parametric in terms of the specification of random utility functions, but operates under the assumption that the stochastic components have an absolutely continuous distribution. A second line of future research might re-examine the propositions of this paper when the demand variables are not continuously distributed, but only take on, e.g., integer values.

A Proofs

A.1 Proof of Lemma 1

Since $\dim(\epsilon) = K$, $\text{rk}(\Sigma) = K$ as well. Suppose $K < J$.

Case (i): Since x is uniformly Lipschitz, there exists a finite scalar κ, such that $\|x(\epsilon) - x(\epsilon')\| \leq \kappa\|\epsilon - \epsilon'\|$, except on a set of measure zero, where $\|\cdot\|$ denotes Euclidean distance; κ is a uniform Lipschitz constant. In particular,

$$\|x(\epsilon) - x(0)\| \leq \kappa\|\epsilon\|.$$

This is equivalent to $(x(\epsilon) - x(0))'(x(\epsilon) - x(0)) \leq \kappa^2\epsilon'\epsilon$, so that

$$\begin{aligned} & \text{tr}((x(\epsilon) - x(0))'(x(\epsilon) - x(0))) \leq \kappa^2\text{tr}(\epsilon'\epsilon) \\ \iff \quad & \text{tr}(\text{var}(x(\epsilon)) + (E[x(\epsilon)] - x(0))\,(E[x(\epsilon)] - x(0))') \leq \kappa^2\text{tr}\,(\Sigma)\,, \end{aligned}$$

where the last inequality follows from the preceding one by commuting the linear operators $\text{tr}(\cdot)$ and $E[\cdot]$. The last inequality then implies that $\text{var}(x(\epsilon))$ exists, i.e. has finite elements, because $p'\,(E[x(\epsilon)] - x(0))\,(E[x(\epsilon)] - x(0))'\,p \geq 0$ for all $p \in \mathbb{R}^J$; the existence of $E[x(\epsilon)]$ follows, of course, for free.

Furthermore, Lipschitz continuity of $x(\epsilon)$ implies that $|x_j(\epsilon_k e_k) - x(0)| \leq \kappa|\epsilon_k|$, for $j = 1, \ldots, J$ and any k, where e_k is the K dimensional unit vector with 1 in position k; similarly, $|x_j(\epsilon + \alpha e_k) - x(\epsilon)| \leq \kappa|\alpha|$ for any scalar α. The triangle inequality, then, implies that $|x_j(\epsilon) - x(0)| \leq \kappa\sum_{k=1}^{K}|\epsilon_k|$ for all j. Hence, there exists a $J\times K$ matrix $\mathbf{K}$, for which $|x(\epsilon) - x(0)| \leq \mathbf{K}|\epsilon|$. Again by the triangle inequality, it follows that $|x(\epsilon) - E[x(\epsilon)]|$ is stochastically dominated by a constant and the random variable $\mathbf{K}|\epsilon|$. If $\mathbf{K}$ has $J > K$, this random variable has a singular distribution, because its covariance matrix has rank at most K. Its support, therefore, is restricted to some subspace of $\mathbb{R}^J$ and its dimension is no larger than K. Since $|x(\epsilon) - E[x(\epsilon)]|$ is bounded by this random variable, it must also lie in this subspace, and so its distribution is degenerate in $\mathbb{R}^J$.

Case (ii): Since the map x is C^1, the rank of the map x in a neighborhood of any ϵ is $\text{rk}(\nabla_\epsilon x(\epsilon)) \leq \min\{K, J\}$. Since $K < J$ by hypothesis, the Rank Theorem[15] implies that x

[15] See, e.g, Boothby, Theorem 7.1, p.47

in a neighborhood of ϵ is at most a K-dimensional submanifold of $\mathbb{R}^J$. This implies that the distribution of $x(\epsilon)$ is degenerate. Note that the moments of ϵ are not needed in this case.

This completes the argument supporting of the claim. □

A.2 Proof of Lemma 2

The condition that the bordered Hessian of U have non-zero determinant for all $\mathbf{x}$ and ϵ, together with the Implicit Function Theorem, implies that $h^M(\mathbf{p}, m, \theta, \epsilon)$ is continuously differentiable, so that the Jacobian $\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)$ exists. Obviously, if $\dim(\epsilon) = J - 1$ and the Jacobian of h^M with respect to ϵ exists and has full column rank with probability one for all $\mathbf{p}$ and m, then $\dim(\mathbf{x}_{-J}) = J - 1$, i.e. the distribution of the endogenous variables is non-degenerate. The strategy is therefore to show that the Jacobian has full column rank.

Define the stochastic expenditure function $e(\mathbf{p}, u, \epsilon) = \min_{\mathbf{x}\in\mathbb{R}^J_+}\{\mathbf{p}'\mathbf{x} : U(\mathbf{x}, \theta, \epsilon) \geq u\}$; for notational simplicity, the dependence on θ is suppressed throughout the remainder of the argument. Then, the system of stochastic Hicksian demand functions is $h^H(\mathbf{p}, u, \epsilon) = \nabla_{\mathbf{p}} e(\mathbf{p}, u, \epsilon)$, and the system of stochastic Marshallian demand functions is $h^M(\mathbf{p}, m, \epsilon) = \arg\max_{\mathbf{x}\in\mathbb{R}^J_+}\{U(\mathbf{x}, \theta, \epsilon) : \mathbf{p}'\mathbf{x} = m\}$, with $h^M(\mathbf{p}, m, \epsilon) = (h(\mathbf{p}, m, \epsilon)', h_J(\mathbf{p}, m, \epsilon))'$. From these definitions, it follows that

$$\begin{aligned} \nabla_\epsilon h^H(\mathbf{p}, u, \epsilon) &= \nabla_\epsilon h^M(\mathbf{p}, m, \epsilon) + \nabla_m h^M(\mathbf{p}, m, \epsilon)\nabla_\epsilon e(\mathbf{p}, u, \epsilon)' \\ &= \nabla_\epsilon h^M(\mathbf{p}, m, \epsilon) + \nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}'\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon), \end{aligned}$$

so that

$$\nabla_\epsilon h^M(\mathbf{p}, m, \epsilon) = [\mathbf{I} - \nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}']\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon). \tag{A-5}$$

Therefore, $\text{rk}(\nabla_\epsilon h^M(\mathbf{p}, m, \epsilon)) \leq \min\{\text{rk}(\mathbf{I} - \nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}'), \text{rk}(\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon))\}$ for each ϵ and $\mathbf{p}, m$.

Consider the first matrix on the right-hand side of ($A - 5$). Note that it follows from Brown and Walker (1989), Theorem 2, that if the demand model exhibits a nonzero income derivative, then this matrix must be stochastic. Since the rank of $\nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}'$ is one for each ϵ, the rank of $\mathbf{I} - \nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}'$ is at most $J - 1$ for each ϵ. In fact, the nullspace associated with the row space of this matrix is one-dimensional, consisting of vectors collinear with $\mathbf{p}$, since $\mathbf{p}'\nabla_m h^M(\mathbf{p}, m, \epsilon) = 1$ w.p.1 from the budget constraint; there are no other, linearly independent vectors in the nullspace since, if there were $\tilde{\mathbf{p}}$, linearly independent of $\mathbf{p}$, with $\tilde{\mathbf{p}}' - \tilde{\mathbf{p}}'\nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}' = \mathbf{0}$, the fact that $\tilde{\mathbf{p}}'\nabla_m h^M(\mathbf{p}, m, \epsilon)$ is a scalar for all ϵ leads to the contradiction that $\tilde{\mathbf{p}}$ is collinear with $\mathbf{p}$. Similarly, the nullspace of the column space is one-dimensional, given by the vectors collinear with $\nabla_m h^M(\mathbf{p}, m, \epsilon)$, for given $\epsilon, \mathbf{p}, m$. Thus, $\text{rk}(\mathbf{I} - \nabla_m h^M(\mathbf{p}, m, \epsilon)\mathbf{p}') = J - 1$ w.p.1.

Now consider the second matrix $\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon)$. Suppose there exists a nonzero vector $d\epsilon$ such that $\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon)d\epsilon = \mathbf{0}$ for some $\mathbf{p}, u$ and on a set S_ϵ of values ϵ with nonzero measure. Then, on S_ϵ,

$$\frac{1}{\delta}\|h^H(\mathbf{p}, u, \epsilon) - h^H(\mathbf{p}, u, \epsilon + \delta d\epsilon)\| \to 0 \quad \text{as } \delta \to 0,$$
$$U(h^H(\mathbf{p}, u, \epsilon), \theta, \epsilon) = U(h^H(\mathbf{p}, u, \epsilon + d\epsilon), \theta, \epsilon + d\epsilon) \equiv u \quad \text{for all } \epsilon.$$

By hypothesis, this holds for some $\mathbf{p}, u$ and $\epsilon \in S_\epsilon \subset \mathbb{R}^{J-1}$, with positive measure $\Pr(S_\epsilon) > 0$. This implies by the definition of Hicksian demand functions that, at u,

$$\begin{aligned} \frac{1}{p_J}\mathbf{p}_{-J} &= \left[\text{MRS}_{jJ}(h^H(\mathbf{p}, u, \epsilon), \epsilon)\right]_{j=1,\dots,J-1} \\ &= \left[\text{MRS}_{jJ}(h^H(\mathbf{p}, u, \epsilon + \delta d\epsilon), \epsilon + \delta d\epsilon)\right]_{j=1,\dots,J-1}, \end{aligned}$$

and therefore, by differentiation with respect to δ,

$$\nabla_{\epsilon'}\text{MRS}_{jJ}(\mathbf{x}, \epsilon)d\epsilon = \mathbf{0} \quad \text{for } j = 1, \dots, J-1$$

on a set of positive ϵ-measure. This contradicts *A1*, by which the Jacobian with respect to ϵ of the J-dimensional vector of stochastic marginal utilities has full column rank almost everywhere, and therefore the $J - 1$-dimensional vector of marginal rates of substitution has a full rank Jacobian. So $\text{rk}(\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon)) = J - 1$ w.p.1.

Therefore, the nullspace associated with the column space of $\nabla_\epsilon h^M(\mathbf{p}, m, \epsilon)$ is given by vectors $\gamma \in \mathbb{R}^{J-1}$ satisfying $\nabla_m h^M(\mathbf{p}, m, \epsilon) = \nabla_\epsilon h^H(\mathbf{p}, u, \epsilon)\gamma$. Since the image of the linear transformation of γ is a one-dimensional space and the transformation matrix $\nabla_\epsilon h^H(\mathbf{p}, u, \epsilon)$ has full rank, the vectors γ belong to a one-dimensional space. Combining these results, $\text{rk}(\nabla_\epsilon h^M(\mathbf{p}, m, \epsilon)) = J - 1$ w.p.1, and so there exists a subset of $J - 1$ rows of $\nabla_\epsilon h^M(\mathbf{p}, m, \epsilon)$, which, w.l.o.g., can be taken to consist of the first $J - 1$ rows, so that $\text{rk}(\nabla_\epsilon h(\mathbf{p}, m, \epsilon)) = J - 1$ w.p.1. This completes the proof of the lemma. □

Remark: (*An Alternative Proof*)

There exists another way to prove sufficiency of $\text{rk}(\nabla_{\mathbf{x}\epsilon}U(\mathbf{x},\theta,\epsilon)) = J-1$ w.p.1 and the condition on the bordered Hessian, i.e. of assumption (*A1*), for $\nabla_{\epsilon}h^M(\mathbf{p},m,\theta,\epsilon) = J-1$ w.p.1. It is essentially an application of the Implicit Function Theorem. Consider the system of equations

$$\begin{bmatrix} \frac{1}{p_J}\mathbf{p}_{-J} \\ m \end{bmatrix} - \begin{bmatrix} \mathbf{MRS}(\mathbf{x},\theta,\epsilon) \\ \mathbf{p}'\mathbf{x} \end{bmatrix} = \mathbf{0} =: M(\mathbf{x},\mathbf{p},m,\theta,\epsilon), \tag{A-6}$$

where $\mathbf{MRS}(\mathbf{x},\theta,\epsilon) = [MRS_{jJ}(\mathbf{x},\theta,\epsilon)]_{j=1,\dots,J-1}$ denotes the vector of $J-1$ stochastic marginal rates of substitution of the $J-1$ inside goods with the outside good. The function M defines an implicit relationship between $\mathbf{x}$ and ϵ, given $\mathbf{p}, m$ and θ. Given any ϵ and θ, this system can be solved for $\mathbf{x}$ as a function of $\mathbf{p}$ and m, since $U \in \mathcal{U}$ implies that the solution for $\mathbf{x}$, $h(\mathbf{p},m,\theta,\epsilon)$, exists and is continuous in $\mathbf{p}$ and m. The condition on the bordered Hessian implies in addition that it is differentiable. In particular, the $J \times J$ matrix $\nabla_{\mathbf{x}}M(\mathbf{p},m,\theta,\epsilon)$ has full rank for all $\epsilon, \mathbf{p}$ and m. Then, for all ϵ, given $\mathbf{p}$ and m,

$$\mathbf{0} = \nabla_{\mathbf{x}}M(\mathbf{x},\mathbf{p},m,\theta,\epsilon)d\mathbf{x} + \nabla_{\epsilon}M(\mathbf{x},\mathbf{p},m,\theta,\epsilon)d\epsilon.$$

From this, it follows that

$$d\mathbf{x} = -\left(\nabla_{\mathbf{x}}M(\mathbf{p},m,\theta,\epsilon)\right)^{-1} \begin{bmatrix} \nabla_{\epsilon}\mathbf{MRS}(\mathbf{x},\theta,\epsilon) \\ \mathbf{0}'_{J-1} \end{bmatrix} d\epsilon = \begin{bmatrix} \nabla_{\epsilon}h^M(\mathbf{p},m,\theta,\epsilon)d\epsilon \\ dx_J \end{bmatrix},$$

where the Jacobian $\nabla_{\epsilon}h^M(\mathbf{p},m,\theta,\epsilon)$ is seen to be decomposed into two matrices, the first of which has rank $J-1$ since $\nabla_{\mathbf{x}}M$ has rank J because $U \in \mathcal{U}$, while the second has full rank $J-1$ by *A10*, so that the Jacobian inherits its rank. This completes the argument. □

A.3 Proof of Lemma 3

Given $(\mathbf{p}',m)' > \mathbf{0}$ and $\theta \in \Theta$, consider any $\bar{\mathbf{x}} \in \left(0, \frac{m}{p_1}\right) \times \ldots \times \left(0, \frac{m}{p_J}\right)$. Since $U \in \mathcal{U}$, $\bar{\mathbf{x}}$ uniquely solves the first-order conditions

$$\begin{aligned} \frac{1}{p_J}\mathbf{p}_{-J} &= \frac{\nabla_{\mathbf{x}_{-J}}U(\bar{\mathbf{x}},\theta,\bar{\epsilon})}{\nabla_{x_J}U(\bar{\mathbf{x}},\theta,\bar{\epsilon})} \\ \mathbf{p}'\bar{\mathbf{x}} &= m, \end{aligned}$$

for some $\bar{\epsilon}$, since *A1a* implies that $\nabla_{\mathbf{x}} U(\mathbf{x}, \theta, \cdot)$ is surjective.

Suppose there exists a neighborhood $E(\bar{\epsilon})$ about $\bar{\epsilon}$ such that $P(E(\bar{\epsilon})) > 0$ and $\forall \epsilon \in E(\bar{\epsilon}) :$ $h^M(\mathbf{p}, m, \theta, \epsilon) = h^M(\mathbf{p}, m, \theta, \bar{\epsilon}) = \bar{\mathbf{x}}$. This implies that

$$\nabla_{\mathbf{x}} U(\bar{\mathbf{x}}, \theta, \epsilon) \propto \nabla_{\mathbf{x}} U(\bar{\mathbf{x}}, \theta, \epsilon) \quad \forall \epsilon \in E(\bar{\epsilon}).$$

Therefore, for any $\epsilon \in E(\bar{\epsilon})$, the vector $\epsilon - \bar{\epsilon}$ is orthogonal to the associated change in the gradient $\nabla_{\mathbf{x}} U(\bar{\mathbf{x}}, \theta, \epsilon) - \nabla_{\mathbf{x}} U(\bar{\mathbf{x}}, \theta, \epsilon)$, or

$$\nabla_{\mathbf{x}\epsilon'} U(\mathbf{x}, \theta, \epsilon)(\epsilon - \bar{\epsilon}) = \mathbf{0} \quad \forall \epsilon \in E(\bar{\epsilon}).$$

Since $P(E(\bar{\epsilon})) > 0$, this contradicts *A1*. Therefore, $h(\mathbf{p}, m, \theta, \cdot)$ is injective almost surely, for all $(\mathbf{p}', m)' > \mathbf{0}$ and θ. This completes the argument. □

A.4 Proof of Proposition 1

It follows from assumptions *A1'* that the inverse with respect to ϵ of the system of demand functions $g(\mathbf{p}, m, \theta, \mathbf{x}_{-J}) = \epsilon$ exists, so that, by assumption *A2*,

$$E[g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] = \mathbf{0} \tag{A-7}$$

$$E[g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J}) g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})'] = \Sigma_0, \tag{A-8}$$

where the expectation is taken with respect to the joint distribution of the vector ϵ, which, by standard axioms about preferences, is independent of $\mathbf{p}$ and m (assumption *A2*). Then, for any $M - (J-1) \times (J-1)$ array of instruments $z(\mathbf{p}, m, \theta)$, with $M \geq K$ and for $\theta \in \Theta$, it follows that also

$$E[z(\mathbf{p}, m, \theta) g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] = \mathbf{0},$$

where the expectation is taken with respect to the joint distribution of $(\mathbf{p}', m, \epsilon')'$. Therefore, Σ_0 is identified from the set of conditional second moment conditions ($A-8$), whether or not θ_0 is identified.[16] The vector θ_0 is identified from the set of unconditional first moment conditions $E[Z(\mathbf{p}, m, \theta) g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] = \mathbf{0}$, where $Z(\mathbf{p}, m, \theta) = [\mathbf{I}_{J-1}, z(\mathbf{p}, m, \theta)']'$ is an $M \times (J-1)$ dimensional array of instruments.

[16] This is analogous to identification of the variance-covariance matrix of the disturbances in the ordinary linear regression model. This matrix is identified even if the covariate matrix does not have full column rank since the conditional mean of the response variables is identified.

In particular, choosing instruments $Z(\mathbf{p}, m, \theta_0)$, θ_0 is locally identified if and only if

$$E[Z(\mathbf{p}, m, \theta_0)g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] \neq E[Z(\mathbf{p}, m, \theta_0)g(\mathbf{p}, m, \tilde{\theta}, \mathbf{x}_{-J}),$$

whenever $\theta_0 \neq \tilde{\theta}$ for any $\tilde{\theta}$ in a neighborhood of θ_0. Thus, for identification of θ_0, it is sufficient that

$$\begin{aligned} & E[Z(\mathbf{p}, m, \theta_0)g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] - E[Z(\mathbf{p}, m, \theta_0)g(\mathbf{p}, m, \tilde{\theta}, \mathbf{x}_{-J})] \\ \approx\; & E[Z(\mathbf{p}, m, \theta_0)\nabla_\theta g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})|_{\theta=\theta_0}]d\theta \\ \neq\; & \mathbf{0}, \end{aligned} \tag{A-9}$$

for $d\theta = \tilde{\theta} - \theta_0$ and any $\tilde{\theta}$ in a neighborhood of θ_0. Thus, local identification of θ_0 follows if the classical rank condition sufficient for identification in Generalized Method of Moments setups is met, namely that $E[Z(\mathbf{p}, m, \theta_0)\nabla_\theta g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})|_{\theta=\theta_0}]$ have constant, full column rank K in a neighborhood of θ_0; see, for example, Newey and McFadden (1994), Lemma 2.3.[17]

To show that the sufficient condition $(A-9)$ is met, consider the identity

$$\mathbf{x}_{-J} = h(\mathbf{p}, m, \theta, g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})),$$

which holds for all $\mathbf{x}_{-J}, \mathbf{p}, m$ and θ. Differentiating with respect to θ,

$$\nabla_\theta h(\mathbf{p}, m, \theta, \epsilon) + \nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)\nabla_\theta g(\mathbf{p}, m, \theta, \mathbf{x}_{-J}) = \mathbf{0},$$

for all x, p, m. Therefore,

$$\begin{aligned} \nabla_\theta g(\mathbf{p}, m, \theta, \mathbf{x}_{-J}) &= -(\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon))^{-1}\nabla_\theta h(\mathbf{p}, m, \theta, \epsilon) \\ &= -[\mathbf{I}_{J-1}, (\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon))^{-1}\nabla_{\theta_2} h(\mathbf{p}, m, \theta, \epsilon)] \\ &= -[\mathbf{I}_{J-1}, d(\mathbf{p}, m, \theta, \epsilon)] \end{aligned}$$

where $d(\mathbf{p}, m, \theta, \epsilon) = (\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon))^{-1}\nabla_{\theta_2} h(\mathbf{p}, m, \theta, \epsilon)$ and the partition on the right-hand side of the last equalities, w.l.o.g., can be achieved by appropriate labeling of the elements of θ. This implies, that, conditional on $\mathbf{p}$ and m, at θ_0,

$$E_\epsilon[Z(\mathbf{p}, m, \theta_0)\nabla_\theta g(\mathbf{p}, m, \theta, \mathbf{x}_{-J})|_{\theta=\theta_0}] = -\begin{bmatrix} \mathbf{I}_{J-1} & D(\mathbf{p}, m, \theta_0) \\ z(\mathbf{p}, m, \theta_0) & z(\mathbf{p}, m, \theta_0)D(\mathbf{p}, m, \theta_0) \end{bmatrix},$$

[17]See also Rothenberg (1971), Theorem 1, and, for identification conditions that imply the ones required by this theorem and cover a more general class of models, Roehrig (1988).

for $D(\mathbf{p}, m, \theta) = E_\epsilon[d(\mathbf{p}, m, \theta, \epsilon)]$. Then, choosing instruments $z(\mathbf{p}, m, \theta_0) = D(\mathbf{p}, m, \theta_0)'$ and taking also expectation with respect to the joint distribution of $(\mathbf{p}', m)'$, the last equality implies

$$\begin{aligned} E[Z(\mathbf{p}, m, \theta_0)\nabla_\theta g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})] &= -\begin{bmatrix} \mathbf{I}_{J-1} & E_{\mathbf{p},m}[D(\mathbf{p}, m, \theta_0)] \\ E_{\mathbf{p},m}[D(\mathbf{p}, m, \theta_0)]' & E_{\mathbf{p},m}[D(\mathbf{p}, m, \theta_0)'D(\mathbf{p}, m, \theta_0)] \end{bmatrix} \\ &= -E_{\mathbf{p},m}\left[E_\epsilon[\nabla_\theta g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})]' E_\epsilon[\nabla_\theta g(\mathbf{p}, m, \theta_0, \mathbf{x}_{-J})]\right]. \end{aligned}$$

Therefore, θ_0 is locally identified if $D(\mathbf{p}, m, \theta)$ has a nonsingular variance-covariance matrix for θ in a neighborhood of θ_0, i.e. if for any nonzero $\alpha \in \mathbb{R}^{K-(J-1)}$,

$$\begin{aligned} 0 &< \alpha' E_{\mathbf{p},m}[D(\mathbf{p}, m, \theta)'D(\mathbf{p}, m, \theta)]\alpha - \alpha' E_{\mathbf{p},m}[D(\mathbf{p}, m, \theta)]' E_{\mathbf{p},m}[D(\mathbf{p}, m, \theta)]\alpha \qquad \text{(A-10)} \\ &= E_{\mathbf{p},m}[\mathcal{D}(\mathbf{p}, m, \theta, \alpha)'\mathcal{D}(\mathbf{p}, m, \theta, \alpha)] - E_{\mathbf{p},m}[\mathcal{D}(\mathbf{p}, m, \theta, \alpha)]' E_{\mathbf{p},m}[\mathcal{D}(\mathbf{p}, m, \theta, \alpha)] \\ &= \operatorname{Var}(\mathcal{D}(\mathbf{p}, m, \theta, \alpha)) \end{aligned}$$

for the vector $\mathcal{D}(\mathbf{p}, m, \theta, \alpha) = D(\mathbf{p}, m, \theta)\alpha$ and θ in a neighborhood of θ_0. Thus, it is sufficient to show that for any nonzero vector $\alpha \in \mathbb{R}^{K-(J-1)}$ and θ in a neighborhood of θ_0, $\mathcal{D}(\mathbf{p}, m, \theta, \alpha)$ has a non-degenerate distribution on $\mathbb{R}^{J-1}$.

The following properties of $d(\mathbf{p}, m, \theta, \epsilon)$ can be shown for any $\theta \in \Theta$:

(i) $d(\mathbf{p}, m, \theta, \epsilon)\alpha \neq 0$ a.s. for all nonzero $\alpha \in \mathbb{R}^{K-(J-1)}$ (by *A1* and $U \in \mathcal{U}$)[18];

[18] Examine the two matrices that build up $d(\mathbf{p}, m, \theta, \epsilon)$. Notice first that $x_j \in [0, \frac{m}{p_j}]$, and therefore

$$\omega_j := x_j - E_\epsilon[h_j(\mathbf{p}, m, \theta, \epsilon)] \in \left[-E_\epsilon[h_j(\mathbf{p}, m, \theta, \epsilon)], \frac{m}{p_j} - E_\epsilon[h_j(\mathbf{p}, m, \theta, \epsilon)]\right],$$

which shows that the conditional residual ω_j has a support that depends on p_j and m and, possibly, all other prices $\mathbf{p}_{-j}$. This implies that $\nabla_\epsilon h(\mathbf{p}, m, \theta\epsilon)$ depends on all prices $\mathbf{p}$ and m. Also, assumption *A1* implies that $\operatorname{rk}(\nabla_\epsilon h(\mathbf{p}, m, \theta, \epsilon)) = J - 1$ for all $\mathbf{p}$, m and ϵ.

The second matrix $\nabla_{\theta_2} h(\mathbf{p}, m, \theta, \epsilon)$ is of dimension $(J-1) \times (K-(J-1))$, where $K-(J-1) \leq J-1$ by *A2*. It is easy to show that its rows are linearly independent with positive probability since otherwise the associated goods would act as perfect substitutes, which violates the assumption of strict quasi-concavity of $U \in \mathcal{U}$. To see this, suppose that $\nabla_{\theta_2} h_j = \alpha_k \nabla_{\theta_2} h_k + \alpha_m \nabla_{\theta_2} h_m$ w.p.1. Then, integrating back and substituting out the proportional terms, $x_j + x_k + x_m = H_{jkm}(\mathbf{p}, m, \theta_1, \epsilon)$, for some function H that summarizes the price and income dependence of the components of h_j, h_k and h_m that do not depend on θ_2.

Combining these results, $\operatorname{rk}(d(\mathbf{p}, m, \theta, \epsilon)) = K - (J-1)$ w.p.1, and therefore $d(\mathbf{p}, m, \theta, \epsilon)\alpha \neq 0$ a.s. for any nonzero $\alpha \in \mathbb{R}^{K-(J-1)}$.

(ii) all elements of $d(\mathbf{p}, m, \theta, \epsilon)$ depend on all prices $\mathbf{p}$, m and ϵ (by the assumption on the stochastic gradient of the utility function, *A2*)[19];

(iii) $\|d(\mathbf{p}, m, \theta, \epsilon)\| < \infty$ a.s. (by *A1*);

(iv) $d(\mathbf{p}, m, \theta, \epsilon)$ is homogeneous of degree zero in $\mathbf{p}$ and m.

These properties imply that

$$\nabla_{\mathbf{p},m}\mathcal{D}(\mathbf{p}, m, \theta, \alpha)\begin{bmatrix}\mathbf{p}\\ m\end{bmatrix} = E_\epsilon\left[\nabla_{\mathbf{p},m}\left(d(\mathbf{p}, m, \theta, \epsilon)\alpha\right)\right]\begin{bmatrix}\mathbf{p}\\ m\end{bmatrix} = \mathbf{0}$$

for all $\mathbf{p}, m$ and α. Since $d(\mathbf{p}, m, \theta, \epsilon)$ depends on all $\mathbf{p}$ and m,

$$\nabla_{\mathbf{p}_{-J},m}\mathcal{D}(\mathbf{p}, m, \theta, \alpha)\begin{bmatrix}\mathbf{p}_{-J}\\ m\end{bmatrix} \neq \mathbf{0},$$

for all $\mathbf{p}, m, \alpha$. By *A3*, therefore, $\mathcal{D}(\mathbf{p}, m, \theta, \alpha)$ has a non-degenerate distribution on $\mathbb{R}^{J-1}$, which implies ($A - 10$). This completes the proof. □

A.5 Proof of Corollary 2

From the first-order conditions,

$$E\left[\frac{1}{p_J}\mathbf{p}_{-J} - \mathbf{MRS}((h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))', \theta, \epsilon)\middle| \mathbf{p}, m\right] = \mathbf{0}.$$

[19]Suppose this was not the case. Then, one of the columns of $\nabla_{\theta_2}h(\mathbf{p}, m, \theta, \epsilon)$ must be proportional to a column in $\nabla_{\theta_1}h(\mathbf{p}, m, \theta, \epsilon)$ for all $\mathbf{p}, m, \epsilon$. From the first order conditions,

$$\left[\nabla_{\mathbf{x}\mathbf{x}'}U(\mathbf{x}, \theta, \epsilon)\nabla_\theta h(\mathbf{p}, m, \theta, \epsilon) + \nabla_{\mathbf{x}\theta}U(\mathbf{x}, \theta, \epsilon)\right]d\theta = \mathbf{p}\left[\frac{\partial\lambda}{\partial\theta}\right]' d\theta,$$

where λ denotes the shadow value of income. If $\nabla_{\theta_{1j}}h(\mathbf{p}, m, \theta, \epsilon) = \delta\nabla_{\theta_{2k}}h(\mathbf{p}, m, \theta, \epsilon)$ for all $\mathbf{p}, m, \epsilon$ and some $j \in \{1, \ldots, J-1\}$, $k \in \{1, \ldots, K-(J-1)\}$ and $\delta \neq 0$, then

$$\nabla_{\mathbf{x}\theta_{1j}}U(\mathbf{x}, \theta, \epsilon) - \delta\nabla_{\mathbf{x}\theta_{2k}}U(\mathbf{x}, \theta, \epsilon) = \mathbf{p}\left[\frac{\partial\lambda}{\partial\theta_{1j}} - \delta\frac{\partial\lambda}{\partial\theta_{2k}}\right],$$

for all $\mathbf{x} \in \mathbb{R}^J_+$. Since the right-hand side depends on $\mathbf{p}$, while the left-hand side does not, the right-hand side must be zero, leading to a contradiction of *A2*.

Therefore, θ is locally identified if

$$E\left[Z(\mathbf{p},m,\theta_0)\nabla_\theta \mathbf{MRS}((h(\mathbf{p},m,\theta,\epsilon)',h_J(\mathbf{p},m,\theta,\epsilon))',\theta,\epsilon)\right]d\theta \neq \mathbf{0},$$

for some choice of instruments $Z(\mathbf{p},m,\theta_0)$, which, as before, can be any function of prices, outlays and optionally the parameters of interest.

Consider the following identities:

$$\begin{aligned} \frac{1}{p_J}\mathbf{p}_{-J} &= \mathbf{MRS}((h(\mathbf{p},m,\theta,\epsilon)',h_J(\mathbf{p},m,\theta,\epsilon))',\theta,\epsilon) \qquad \forall \ \ \epsilon,\mathbf{p},m,\theta, \\ \mathbf{x}_{-J} &= h(\mathbf{p},m,\theta,g(\mathbf{p},m,\theta,\mathbf{x}_{-J})) \qquad \forall \ \ \mathbf{x}_{-J},\mathbf{p},m,\theta. \end{aligned}$$

From the first, it follows that, for any $d\theta \neq \mathbf{0}$,

$$\left(\nabla_\theta \mathbf{MRS}(\mathbf{x},\theta,\epsilon) + \nabla_\mathbf{x}\mathbf{MRS}(\mathbf{x},\theta,\epsilon)\nabla_\theta h + \nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta,\epsilon)\nabla_\theta g\right)d\theta = \mathbf{0},$$

where $\nabla_\theta h = \nabla_\theta h(\mathbf{p},m,\theta,\epsilon)$ and $\nabla_\theta g = \nabla_\theta g(\mathbf{p},m,\theta,\mathbf{x}_{-J})$, all expressions are evaluated at $\mathbf{x} = (h(\mathbf{p},m,\theta,\epsilon)',h_J(\mathbf{p},m,\theta,\epsilon))'$, for any $\mathbf{p},m,\theta$ and ϵ. Then, from the second identity and the budget constraint,

$$\nabla_\theta \mathbf{MRS}(\mathbf{x},\theta,\epsilon) = \left(\nabla_\mathbf{x}\mathbf{MRS}(\mathbf{x},\theta,\epsilon)\begin{bmatrix} \nabla_\epsilon h \\ -\frac{1}{p_J}\mathbf{p}'_{-J}\nabla_\epsilon h \end{bmatrix} - \nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta,\epsilon)\right)\nabla_\theta g, \qquad \text{(A-11)}$$

at $\mathbf{x} = (h(\mathbf{p},m,\theta,\epsilon)',h_J(\mathbf{p},m,\theta,\epsilon))'$, a $(J-1)\times K$-dimensional matrix equation.

Recall from the proof of Proposition 1 that it also follows from the second identity that

$$\nabla_\theta g(\mathbf{p},m,\theta,\mathbf{x}_{-J}) = -\left[\mathbf{I}_{J-1}, d(\mathbf{p},m,\theta,\epsilon)\right],$$

where $d(\mathbf{p},m,\theta,\epsilon) = \left(\nabla_\epsilon h(\mathbf{p},m,\theta,\epsilon)\right)^{-1}\nabla_{\theta_2}h(\mathbf{p},m,\theta,\epsilon)$, by appropriate labeling of the components of θ. Since, by *A2*, $\nabla_{\theta_1}\mathbf{MRS}(\mathbf{x},\theta,\epsilon) = \nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta,\epsilon)$, the matrix system ($A-11$), via its first $J-1$ columns, implies for the $(J-1)\times(J-1)$-dimensional matrix

$$\nabla_\mathbf{x}\mathbf{MRS}(\mathbf{x},\theta,\epsilon)\begin{bmatrix} \nabla_\epsilon h(\mathbf{p},m,\theta,\epsilon) \\ -\frac{1}{p_J}\mathbf{p}'_{-J}\nabla_\epsilon h(\mathbf{p},m,\theta,\epsilon) \end{bmatrix} = \mathbf{0}, \qquad \forall \ \ \mathbf{x},\mathbf{p},m,\theta,\epsilon.$$

This implies that

$$\begin{aligned} \nabla_\theta \mathbf{MRS}(\mathbf{x},\theta,\epsilon) &= \nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta,\epsilon)\left[\mathbf{I}_{J-1}, d(\mathbf{p},m,\theta,\epsilon)\right] \\ &= -\nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta,\epsilon)\nabla_\theta g(\mathbf{p},m,\theta,\mathbf{x}_{-J}) \end{aligned}$$

again evaluated at $\mathbf{x} = (h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))'$.

The remainder of the proof proceeds in the same spirit as the proof of Proposition 1. That is, it is shown that $E\left[\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon)\right]$ and $E\left[\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon)\right]$ have nonsingular variance-covariance matrices. Then, these can be used to form optimal instruments $Z(\mathbf{p}, m, \theta_0)$, so that the sufficient condition for local identification of θ_0 is met.

From Lemma 1, it follows that $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon)$ has full rank for almost all ϵ. Therefore, $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon) \neq \mathbf{0}$ w.p.1. Without loss of generality, the representative in the equivalence class of random utilities is considered for which $\frac{d}{dx_J} U(\mathbf{x}, \theta, \epsilon)$ does not depends on ϵ. Then, $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) = \left(\frac{d}{dx_J} U(\mathbf{x}, \theta, \epsilon)\right)^{-1} \nabla_{\mathbf{x}_{-J}\epsilon} U(\mathbf{x}, \theta, \epsilon)$. Since strict quasi-concavity of $U \in \mathcal{U}$ precludes constant marginal utilities, it must be that $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon)$ indeed depends on $\mathbf{x}$. This implies that $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon)$, when evaluated at $\mathbf{x} = (h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))'$, depends on all prices and is homogeneous of degree zero in $\mathbf{p}$ and m. Strict monotonicity of $U \in \mathcal{U}$ implies that $(h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))' > \mathbf{0}$, and therefore $\|\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon_j)\| < \infty$ a.s., for $j = 1, \ldots, J-1$, when evaluated at $\mathbf{x} = (h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))'$. Therefore, by the same logic as it was shown before in the proof of Proposition 1 that properties (i)-(iv) of $d(\mathbf{p}, m, \theta, \epsilon)$ imply that $E\left[d(\mathbf{p}, m, \theta, \epsilon) | \mathbf{p}, m\right]$ has a nonsingular variance-covariance matrix, it follows that $E[\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon)]$, evaluated at $\mathbf{x} = (h(\mathbf{p}, m, \theta, \epsilon)', h_J(\mathbf{p}, m, \theta, \epsilon))'$, has a nonsingular variance-covariance matrix.

Similarly, from these properties of $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon)$ and the properities of $d(\mathbf{p}, m, \theta, \epsilon)$, one can derive the analogously that, for any θ,

(i) $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon) \alpha \neq \mathbf{0}$ a.s. for all nonzero $\alpha \in \mathbb{R}^{K-(J-1)}$ (by *A1* and $U \in \mathcal{U}$);

(ii) the elements of $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon) \neq \mathbf{0}$ depend on all elements of $\mathbf{p}, m$ and ϵ (by *A2*);

(iii) $\|\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon)\| < \infty$ a.s. (by *A1*);

(iv) $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon) \neq \mathbf{0}$ is homogeneous of degree zero in $\mathbf{p}$ and m.

From these properties of the matrix product $\nabla_\epsilon \mathbf{MRS}(\mathbf{x}, \theta, \epsilon) d(\mathbf{p}, m, \theta, \epsilon)$ it follows that, if multiplied by any nonzero $\alpha \in \mathbb{R}^{K-(J-1)}$, the resulting vector has a non-degenerate distribution on $\mathbb{R}^{J-1}$. Thus, $E\left[\nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta,\epsilon)D(\mathbf{p},m,\theta,\epsilon)|\mathbf{p},m\right]$ has a nonsingular variance-covariance matrix for any θ.

Therefore, choose the array of optimal, though in practice infeasible instruments

$$Z(\mathbf{p},m,\theta_0) = \left[E\left[\nabla_\epsilon \mathbf{MRS}(\mathbf{x},\theta_0,\epsilon)|\mathbf{p},m\right]', D(\mathbf{p},m,\theta_0)'\right]',$$

where the matrix $\nabla_\epsilon \mathbf{MRS}$ is evaluated at $\mathbf{x} = (h(\mathbf{p},m,\theta_0,\epsilon)', h_J(\mathbf{p},m,\theta_0,\epsilon))'$, and as before $D(\mathbf{p},m,\theta_0) = E\left[d(\mathbf{p},m,\theta_0,\epsilon)|\mathbf{p},m\right]$. The two preceding paragraphs establish that this array of instruments has a non-singular variance-covariance matrix, i.e. that it is of rank K. Therefore, θ is locally identified from the first-order conditions. $\square$.

References

[1] Attanasio, O.P., and G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation", *Review of Economic Studies*, vol.60(3), p.632-649

[2] Aumann, R.J. (1969):"Measurable utility and the measurable choice theorem", in: *La décision*, Editions du Centre National de la Recherche Scientifique, p.15-26

[3] Barten, A.P. (1964): "Consumer Demand Functions under Conditions of Almost Additive Preferences", *Econometrica*, vol.32, p.1-38

[4] Barten, A.P. (1968): "Estimating Demand Equations", *Econometrica*, vol.36(2), p.213-251

[5] Beckert, W (1999): "Estimation of Stochastic Preferences: An Empirical Analysis of Demand for Internet Services", mimeo, Department of Economics, University of California, Berkeley

[6] Boothby, W (1986): *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Orlando: Academic Press

[7] Brown, B.W. and M.B. Walker (1989): "The Random Utility Hypothesis and Inferences in Demand Systems", *Econometrica*, vol.57(4), p.815-829

[8] Brown, D.J. and R.L. Matzkin (1995): "Estimation of a Random Utility Model from Data on Consumer Demand", mimeo, Yale University and Northwestern University

[9] Brown, D.J. and R.L. Matzkin (1995A): "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand", mimeo

[10] Darrough, M., R.A. Pollak and T.J. Wales (1983): "Dynamic and stochastic structure: An analysis of three time series of household budget studies", *The Revies of Economics and Statistics*, vol. 65, p.247-81

[11] Debreu, G. (1972): "Smooth Preferences", *Econometrica*, vol.40(4), p.603-615

[12] Dubin, J.A. and D.L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", *Econometrica*, vol.52(2), p.345 -362

[13] Fuss, M., D.L. McFadden and Y. Mundlak (1978): "A Survey of Functional Forms in the Economic Analysis of Production," in *Production economics : a dual approach to theory and applications*, Melvyn Fuss and Daniel McFadden, eds. Amsterdam: North-Holland

[14] Kapteyn, A., van de Geer, S., van de Stadt, H. and T. Wandbeck (1997): "Interdependent Preferences: An Econometric Analysis", *Journal of Applied Econometrics*, vol.12(6), p. 665-86

[15] Klein, L.R. and H. Rubin (1947): "A Constant-Utility Index of the Cost of Living", *The Revies of Economics and Statistics*, vol. 15(2), No.38, p.84-87

[16] Lewbel, A. (1996): "Demand Systems With and Without Errors: Reconciling Econometric, Random Utility and GARP Models", mimeo, Brandeis University

[17] Mas-Colell, Andreu (1977): "The Recoverability of Consumers' Preferences from Market Demand Behavior", *Econometrica*, vol.45(6), p.1409-1430

[18] McElroy, M.B. (1987): "Additive General Error Models for Production, Cost and Derived Demand and Share Systems", *Journal of Political Economy*, vol.95, p.737-757

[19] McFadden, D.L. (1974): "Conditional logit analysis of qualitative choice behavior", in: *Frontiers in Econometrics*, Zarembka, P., ed., New York: Academic Press

[20] McFadden, D.L. (1981): "Econometric models of probabilistic choice", in: *Structural Analysis of Discrete Data with Econometric Applications*, Manski, C. and D.L. McFadden, eds., Cambridge, Mass. : MIT Press

[21] McFadden, D.L. and K. Train (1998): "Mixed MNL Models for Discrete Response", mimeo, Department of Economics, U.C. Berkeley

[22] Newey, W.K. and D.L. McFadden: "Large Sample Estimation and Hypothesis Testing", in: R.Engle and D.L.McFadden, eds., *Handbook of Econometrics, Vol.4*, Amsterdam: North-Holland

[23] Newey, W.K., J.L Powell and F.Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica*, vol.67(3), p.565-603

[24] Pollak, R.A. and T.J. Wales (1969): "Estimation of the Linear Expenditure System", *Econometrica*, vol.37, p.611-28

[25] Roehrig, Ch. (1988): "Conditions for Identification in Nonparametric and Parametric Models", *Econometrica*, vol.56(2), p.433-447

[26] Rothenberg, T.J. (1971): "Identification in Parametric Models", *Econometrica*, vol.39(3), p.577-591

[27] Ruud, P.A. (1999): *An Introduction to Classical Econometric Theory*, forthcoming in Oxford University Press

[28] Stone, R. (1954): "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand", *Economic Journal*, vol.64, p.511-27

[29] Theil, H. (1965): "The Information Approach to Demand Analysis", *Econometrica*, vol.33, p.67-87

[30] Theil, H. (1967): *Economics and Information Theory*, Amsterdam: North-Holland

[31] Thurstone, L.L. (1927): "A law of comparative judgment", *Psychological Review*, vol.34, p.273-86

[32] Wesley, R. (1976): "Elementary proof of the measurable utility theorem", in: o. Moeschlin and S. Pallaschke, eds., *Game theory and real topics*, Amsterdam: North-Holland

[33] Wieczorek, A. (1980): "On the measurable utility theorem", *Journal of Mathematical economics*, vol.7, p.165-173