Identification in multivariate nonseparable triangular models with applications to Engel curves, the BLP-, and Hedonic models

Florian F Gunsilius*

Brown University

This version: February 7, 2018
Comments welcome.

Abstract

In this article we introduce general nonparametric identification results for nonseparable triangular models with a multivariate first- and second stage. In our main result we prove point-identification of these models, generalizing results from the literature which all require a univariate first or second stage. This enables us to prove nonparametric identification results for the BLP model without the need to make index restriction as well as Hedonic models with multivariate heterogeneity and endogenous characteristics. In addition, we provide a general control variable approach, showing that random variables with a wide array of distributions, even multivariate under a normalization assumption, can be used as a control variable. Using this we provide a fully nonparametric estimation of quantile structural functions of Engel curves, whilst allowing for more than one endogenous variable in the model.

1 Introduction

Over the last two decades several approaches towards identification of nonseparable triangular models of the form

\[ Y = m(X, \varepsilon) \]
\[ X = h(Z, U), \]

have been developed, where \( Y \) is the outcome, \( X \) is an endogenous regressor, \( U \) and \( V \) are latent error terms, and \( m \) and \( h \) are unknown production functions. Some results focus on point-identification of the second stage (d’Haultfoeuille & Février 2015, Torgovitsky 2015a) while others

*Correspondence: Florian_Gunsilius@brown.edu. This is an extended version of my third year paper at Brown University. I thank Susanne Schennach, Ken Chay, Adam McCloskey, Andriy Norets, and Jesse Shapiro for helpful comments. All errors are mine.

All of these results aiming for point-identification of the second stage require a univariate and strictly increasing first-or second stage (in particular Imbens & Newey 2009, d'Haultfoeuille & Février 2015, Torgovitsky 2015a), which limits their practical applicability in settings with general multivariate heterogeneity like the Hedonic model or the BLP model (Matzkin 2007a and Berry & Haile 2014, p. 1754). A generalization of identification results to a multivariate setting without strong artificial functional form assumptions is hence important, in particular for bridging the gap between Economic and Econometric theory.

As the main contribution of this article, we therefore provide a new framework for identification in nonseparable triangular models, which allows us to prove point-identification in nonseparable triangular models while allowing for both first- and second stage to be multivariate, generalizing the point-identification results in Torgovitsky (2015a) and d’Haultfoeuille & Février (2015). In addition, we provide a generalization of the seminal control variable approach (Blundell & Powell 2003, Imbens & Newey 2009). Both theoretical results have a wide range of applications.

As the main application of our main result we obtain the first proof of identification of multi-market Hedonic models with endogenous characteristics, complementing and building on the existing results in Ekeland, Heckman & Nesheim (2004), Heckman, Matzkin & Nesheim (2010), and Chernozhukov, Galichon, Henry & Pass (2014) who all consider Hedonic models with exogenous characteristics. In particular, we give an answer to the open question in the latter article, asking under which conditions one can nonparametrically identify Hedonic models when observable characteristics are endogenous. In a second application, we also show that the BLP model (Berry, Levinsohn & Pakes 1995) can be nonparametrically identified without the need to assume that individual heterogeneity can be captured by an index, complementing the seminal result from Berry & Haile (2014).

As a second theoretical result we prove that variables with a wide variety of distributions can serve as control variables in the case where the endogenous variable is univariate, generalizing Imbens & Newey (2009) who construct a uniformly distributed control variable. This generalization is important, because the “large support assumption” needed for identification can be more easily satisfied for control variables with other distributions than the uniform, leading to tighter identification bounds or even point-identification of average effects in cases where this was not possible previously. In addition, we provide new assumptions under which we can allow for a multivariate first stage $X = h(z,U)$. We apply the results to the 1996 UK Expenditure survey to show that the large support assumption is easier to satisfy for $V$ with a different distribution than the univariate, leading to tighter bounds for identification. Moreover, we estimate quantile structural functions nonparametrically while for the first time allowing for multivariate endogenous variables, considering total family expenditure and children’s age in the household to be endogenous.

The article is structured as follows. In section 2 give a brief overview of the current state of the
literature and contrast our approach to other existing approaches. Section 3 contains the main results: the general identification result for nonseparable triangular models as well as the general control variable approach. Section 4 contains three applications of the main results: a proof of identification for Hedonic models with endogenous characteristics and multivariate heterogeneity, a proof of identification of the BLP model, as well as an application of the general control variable approach to estimating Engel curves. Section 5 concludes. The appendix contains all proofs for the results in the main paper and a brief overview of the Monge-Kantorovich problem from optimal transport and the Brenier map.

2 The literature on identification in nonseparable models

In this section we link our result to the literature on nonseparable models in general and nonseparable triangular models in particular while giving an intuitive overview of the standard assumptions in the literature.

Nonseparable triangular models are an extension of nonseparable models of the form \( Y = m(X, \varepsilon) \) with exogenous \( X \). The literature on identification in these models is large, with several authors seeking to point-identify the production function \( m \) (Matzkin 2003, Matzkin 2007b, Imbens 2007 and references therein) while others predominantly aim for identification of average or marginal effects (Heckman & Vytlacil 2005, Altonji & Matzkin 2005, Höderlein & Mammen 2007, Chernozhukov, Imbens & Newey 2007, Höderlein & Mammen 2009). The literature on identification of such models has been growing ever since and has been fruitfully applied to and extended in different scenarios like single-market Hedonic models (Ekeland, Heckman & Nesheim 2004, Heckman, Matzkin & Nesheim 2010, Chernozhukov, Galichon, Henry & Pass 2014) or non-linear Difference-in-Difference models (Athey & Imbens 2006, d’Haultfoeuille, Höderlein & Sasaki 2013). These results in turn also give rise to many identification results for simultaneous equation models (Matzkin 2008, Berry & Haile 2013, Blundell & Matzkin 2014, Matzkin 2015 and references therein).

All of these results in one way or another require invertibility of the production function \( m \); the standard approach in the majority of approaches is the assumption of strictly increasing and continuous \( m \) which in turn also requires that \( Y \) and \( \varepsilon \) are univariate. In this article we argue that while monotonicity is often a reasonable assumption to make, there is a more general analogue of a strictly monotonically increasing and continuous function which allows for point-identification of the production function in more general settings: measure-preservation of \( m \). In fact, we show that the combination of a unique nonparametric structure in combination with measure preservation of \( m \) leads to identification results which are the natural extensions of the of the identification results which rely on strictly increasing and continuous production functions. Our focus thereby is the setting of nonseparable triangular models.

Allowing for endogenous \( X \) is an important extension of the model, especially for practical purposes as many models of interest take this form. For instance, the BLP model (Berry, Levinsohn
and Hedonic models with multivariate heterogeneity and endogenous characteristics, which are the two applications of our main result; our result is therefore important as the point-identification results in [Torgovitsky (2015a) and d'Haultfoeuille & Fevrier (2015)] cannot be applied in these settings as they require a univariate first- and second stage. Our identification result has other possible applications, for example screening models with multidimensional consumer heterogeneity as in Aryal (2017); the latter introduces an identification result for screening models based on d'Haultfoeuille & Fevrier (2015) and requires strong high level assumptions on the observable distributions as well as strong functional form assumptions on the second stage.

In fact, our general identification result covers four important general aspects which have not or only partially been dealt with in the literature before. First, it is the only result for multivariate nonseparable triangular models, and can be applied to a wide variety of settings. Second, we can allow for the most general functional forms on the production functions $m$ and $h$, without being forced to require monotonicity as Torgovitsky (2015a), d’Haultfoeuille & Fevrier (2015), or Aryal (2017). Third, we prove a new mathematical result which leads to easy-to-check sufficient conditions for identification, generalizing the existing sequencing arguments in Torgovitsky (2015a) and d’Haultfoeuille & Fevrier (2015). Fourth, in our result we can allow for a continuous or discrete (or even binary) instrument $Z$, which can be of lower dimension than $X$, analogous to existing approaches; in our setting, we can allow for lower-dimensional $Z$ even when $m$ and $h$ are truly multivariate functions, and not element-wise monotonic.

Our other theoretical result, the generalization of the control variable approach (Blundell & Powell 2003, Imbens & Newey 2009) can be applied to a wide array of settings. In particular, one can use control variables whenever one does not want to make assumptions on the latent variable $U$. Intuitively, the control variable $V$ mimics the behavior of $U$, in the sense that conditioning on $U$ is the same as conditioning on $V$. In fact, we can actually use the control variable in our general identification result, so that we do not have to make assumptions on $U$. Let us now turn to these results right after introducing the main notation.

**Notation** The standard measurable space is defined by $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ where $\mathcal{B}_{\mathbb{R}^d}$ is the Borel $\sigma$-algebra. For a random variable $X : \Omega \to \mathbb{R}^d$ in some measure space $(\Omega, \mathcal{A}, P)$, the measure $P_X$ is the pushforward measure of $P$ via $X$, i.e. $P_X(E) = P(X^{-1}(E))$ for every Borel set $E \in \mathcal{B}_{\mathbb{R}^d}$. The corresponding distribution function $F_X : \mathbb{R}^d \to [0, 1]$ is defined by $F_X(x) = P_X(X \leq x)$. The support of $X$ is $\mathcal{X}$. Conditional distributions are defined by $F_{X|Z=z,W=w}$ with support $\mathcal{X}_{zw}$. The standard partial order on $\mathbb{R}^d$ induced by the positive cone for $x, x' \in \mathbb{R}^d$ is defined as $x \leq x'$ if and only if $x_i \leq x'_i$ for $i = 1, \ldots, d$. Based on this the distribution function $F_X$ is strictly increasing if $F_{X|Z=z}(x) < F_{X|Z=z}(x')$ whenever $x < x'$ in the standard partial order. Throughout, whenever we require the functions $X = h(Z, U)$ and $Y = m(X, \varepsilon)$ to be invertible, we always mean invertibility with respect to the second argument, i.e. between $X$ and $U$ as well as $Y$ and $\varepsilon$, never with respect to $Z$. Also, we assume $\mathcal{Y} \subseteq \mathbb{R}^d$, $\mathcal{X}, U \subseteq \mathbb{R}^k$, and $Z \subseteq \mathbb{R}^m$. 


3 Main results: Identification in multivariate nonseparable triangular models

This section contains the two main identification results. Contrary to our outline in the first two sections above, we start with the control function approach here, which is section 3.1. After that we introduce the main result of this article in section 3.2, which is the general identification result for nonseparable triangular models. Right now, let us give a general outline of the underlying ideas for both results.

We work within the following model throughout:

\[ Y = m(X, \varepsilon) \]
\[ X = h(Z, U), \]

where the covariate \( X \) is endogenous, i.e. depends on the unobservable error term \( \varepsilon \). \( U \) is the unobservable and independent error term in the first stage. \( m \) and \( h \) are unobservable. The variable \( Z \) is an instrument and fulfills Assumption 1 (Instrument). \( Z \) is an admissible instrument for \( X \), that is (i) \( Z \) is independent of \( U \) and \( \varepsilon \), denoted as \( (\varepsilon, U) \perp \perp Z \), and (ii) \( Z \) has a nonzero effect on \( X \) in the sense that \( F_{X|Z=z} \) is different from \( F_{X|Z=z'} \) for at least one \( x \) in the support \( X \) for all \( z, z' \in Z \).

We allow for the whole or part of the vector \( X \in \mathbb{R}^k \) to be endogenous. In the case where only a part of \( X \) is endogenous, it is customary to write the second stage of (1) as \( Y = m(X, W, \varepsilon) \), where \( W \) are the exogenous covariates, and condition all results on \( W \). For the sake of conciseness we consider all elements of \( X \) endogenous and suppress \( W \) throughout.

The idea to allow for multivariate first- and second stages is to find a sufficiently general functional requirement for \( m \) and \( h \), which seminal results like Imbens & Newey (2009), Torgovitsky (2015a), and d’Haultfœuille & Février (2015) require to be strictly increasing and continuous. The issue is that there is no complete order in higher dimensions, so that those classical identification results based on Matzkin (2003) are not applicable in this setting. We therefore argue that the appropriate generalization of a strictly increasing and continuous function in these settings is a measure preserving isomorphism.

**Definition 1.** A map \( T : \mathcal{E} \rightarrow \mathcal{Y} \) transporting a probability measure \( P_\varepsilon \) onto another probability measure \( P_Y \) is measure preserving if it is measurable \(^1\) and

\[ P_Y(E) = P_\varepsilon(T^{-1}(E)) \]

for every set \( E \) in the Borel \( \sigma \)-algebra \( \mathcal{B}_{\mathbb{R}^d} \) corresponding to \( \mathcal{Y} \). If \( T \) is invertible and its inverse is also measure preserving, it is called a measure-preserving isomorphism.

\(^1\) Measurability of \( T \) means that \( \mathcal{B}_{\mathbb{R}^d} = T^{-1}(\mathcal{B}_{\mathbb{R}^d}) \). \( \mathcal{E} \) denotes the support of \( \varepsilon \).

\(^2\) \( T^{-1}(E) \) denotes the set of points \( e \in \mathcal{E} \) such that \( T(e) \in E \).
A way to check whether a transformation in $\mathbb{R}^d$ is measure preserving is by checking this property for all half-open rectangles of the form $(a,b]$, $a,b \in \mathbb{R}^d$.

**Proposition 1.** $T : \mathcal{X} \subseteq \mathbb{R}^m \to \mathcal{Y} \subseteq \mathbb{R}^n$ with $y = T(x)$ transporting $P_X$ onto $P_Y$ is measure preserving if and only if

$$P_Y((a,b]) = P_X(T^{-1}((a,b])).$$

(3)

**Proof.** This immediately follows from Theorem A.8 in Einsiedler & Ward (2013) and the fact that all rectangles of the form $(a,b]$ form a semi-ring in $\mathcal{B}_Y$ and $\mathcal{B}_X$. \hfill $\square$

The set of all measure preserving isomorphisms between two distribution functions is large. In particular, notice that a strictly increasing and continuous $T$ in the univariate case must be measure preserving since

$$P_Y((-\infty,y]) \equiv P(Y \leq y) = P(T(X) \leq y) = P(X \leq T^{-1}(y)) \equiv P_X(T^{-1}((-\infty,y]),$$

(4)

where the third equality follows from the fact that $T$ is continuous and strictly increasing. Strictly increasing and continuous functions are special measure preserving maps because they map every interval of the form $(-\infty,y_1]$ to an interval of the same form $T^{-1}((-\infty,y_1]) \equiv (-\infty,x_1]$ such that both intervals have the same probability (Figure 1). This requirement makes strictly increasing and continuous functions unique in the class of measure preserving isomorphisms between two fixed distributions $F_X$ and $F_Y$, a property which has first been exploited in Matzkin (2003). On the other hand, measure preservation only requires that an interval $(-\infty,y_1]$ gets mapped to some combination of intervals, a much weaker restriction (Figure 2).

To make these concepts more intuitive consider $\mathcal{X}$ as well as $\mathcal{Y}$ to be a continuum of individuals, respectively. The probability measure $P_X$ then gives the “size” of each (Borel-) subset $E \subseteq \mathcal{X}$ of individuals, and analogous for $P_Y$. The classical idea for identification using monotonicity is

---

$^3$Half-open rectangles in $\mathbb{R}^d$ are the $d$-fold cartesian product of half-open intervals, i.e. $(a,b] := \times_{i=1}^d(a_i,b_i)$, $a_i,b_i \in \mathbb{R}$ and $a = (a_1,\ldots,a_d)'$, $b = (b_1,\ldots,b_d)'$.

$^4$A strictly increasing and continuous function is invertible.

$^5$Completely formally, the image of a measure preserving set need not even be an interval, but could be more general, like a Cantor set for example.
that—in one dimension—requiring the production function \( f : \mathcal{X} \to \mathcal{Y} \) to be strictly increasing and continuous means that \( f \) preserves the ordering of individuals when mapping from \( X \) to \( Y \). This immediately implies that the map \( f \) preserves measure, as a group of people \( E \subset \mathcal{X} \) with size \( P_X(E) \) gets mapped to a group of people \( f(E) \subset \mathcal{Y} \) of size \( P_Y(f(E)) = P_X(E) \), simply by the fact that the order of individuals needs to be preserved, which follows from (4).

Since a strictly increasing and continuous \( f \) is invertible, this is our characterization of measure preservation. As will become clear throughout the two main results in the next two subsections, the main requirement for general identification is measure preservation; if \( m \) is to be point-identified, we in addition need to require \( m \) to be unique. Uniqueness comes from functional form restrictions like (multivariate generalizations of) monotonicity, but monotonicity is simply a sufficient condition for measure preservation and uniqueness, which in turn are a sufficient condition for point-identification of \( m \) as we show in section 3.2.

3.1 Control variable approach

We now generalize the seminal control variable approach from Imbens & Newey (2009) for the non-separable triangular model (1). Blundell & Powell (2003) and Imbens & Newey (2009) introduced the concept of a control variable.

**Definition 2 (Control variable).** A random variable \( V \) is a control variable if \( X \perp \epsilon | V \).

The underlying idea for identification using control variables is as follows. Fix \( U \) by conditioning on some \( u \in U \). Then \( X = h(Z, u) \) is just a function of the exogenous \( Z \), so that by

\[ P_X(T^{-1}((a, b])) \quad \text{and} \quad P_Y((a, b]) \]

\[ T^{-1}((c, d)) \quad \text{and} \quad T^{-1}((a, b]) \]

\[ a, b, c, d \]

Figure 2: General measure preservation

---

6It seems like a tautology that we require uniqueness in order to obtain point-identification. Note, however, that monotonicity and continuity are also just structural assumptions on \( m \) which ensure uniqueness of \( m \). These assumptions are “nonparametric” because in order to check them one would have to check the function at every real (or in the case of continuity: rational) number. Parametric structural assumptions on the other hand need to be checked only at finitely many points. It is in this sense that all these identification results mentioned above are nonparametric, but identification always requires some structural assumptions on the functions, either parametric or nonparametric. So what we are requiring simply is some nonparametric structural assumption which guarantees that the function \( m \) is unique under this, which need not be monotonicity. There are plenty of these structural assumptions. In fact, one general class of functions comes from the theory of optimal transport as we show in section 3.2.
varying $Z$ one can obtain the exogenous effect of $X$ on $Y$ through $Z$, so that average or marginal effects of $m$ will be identifiable. $U$ is unobservable, however. This is where the control function $V$ “mimics” the behavior of $U$ in the sense that conditioning on $V$ is equivalent to conditioning on $U$, so that conditional on $V$ one can obtain the exogenous effect of $X$ on $Y$ by varying $Z$.

Consider the diagram in Figure 3.

\[ P_{X|Z} = g^{-1}(\cdot, z) \]
\[ P_{U} = h^{-1}(\cdot, z) \]

Figure 3: Construction of the control variable $V$

The original result in Imbens & Newey (2009) requires $U$ and $X$ to be univariate and $h(Z, U)$ to be strictly increasing and continuous in $U$ and $g^{-1}(X, Z)$ to be the transform $F_{X|Z}(X, Z)$, so that $U = h^{-1}(X, Z)$ and $F_{X|Z}(x, z) = F_U(h^{-1}(X, Z))$. Then Imbens & Newey (2009) form the control variable $V$ as

\[ V = F_{X|Z}(X, Z) = F_U(h^{-1}(X, Z)) = F_U(U), \]  

where the second equality follows from the fact that $h$ is strictly increasing and continuous in $U$. They then assume $F_U : U \to [0, 1]$ to be strictly increasing and continuous, which makes $F_U : U \to V$ a measurable map whose inverse is also measurable. This implies that the $\sigma$-algebras of $U$ and $V$ coincide, so that conditioning on one is the same as conditioning on the other. Because of (5), $V$ is uniformly distributed on the unit interval $[0, 1]$ by construction. This is rather limiting, as in order to identify average or marginal effects for the second stage, one needs to make the following strong “large support” assumption.

**Assumption 2** (Large support). *The conditional support $V_x$ of the control variable $V$ coincides with the unconditional support $V$ for all $x \in \mathcal{X}$.***

This assumption allows one to integrate out the control variable in order to identify average effects as in the following

**Proposition 2** (Identification of average effects). *Let Assumptions 1 and 2 hold for model (1) and let $V$ be a control variable. If $m$ maps $\varepsilon$ from some (possibly infinite dimensional) metric space $\mathcal{E}$ onto $Y \in \mathbb{R}^d$, then for any integrable function $\Lambda(y)$,

\[ \int \Lambda(m(x, e)) dF_\varepsilon(e) = E(E(\Lambda(Y)|X = x, V)). \]
Proof. It holds that

\[ E(E(\Lambda(Y)\mid X = x, V)) = \int E(\Lambda(Y)\mid X = x, V = v)F_V(dv) \]
\[ = \int_{V_x} \int \Lambda(m(x, e)) F_{\epsilon|X,Y}(de, x, v) F_V(dv) \]
\[ = \int_{V} \int \Lambda(m(x, e)) F_{\epsilon|V}(de, v) F_V(dv) \]
\[ = \int \Lambda(m(x, e)) F_{\epsilon}(de), \]

where the second equality follows from the fact that \( Y = m(x, \epsilon) \) and the third and fourth follow from Definition 2 and the large support assumption which lets us integrate out \( V \).

Proposition 2 is perfectly analogous to the result in Imbens & Newey (2009) and proves identification of average effects of \( m \) if \( V \) is a control variable. The problem with forcing the control variable to follow the uniform distribution on the unit interval is that in practice Assumption 2 is often badly violated, as in the empirical application considered in Imbens & Newey (2009), where the model is used to estimate quantile structural functions of Engel curves from the 1995 UK Family expenditure survey, substituting \( \Lambda(Y) := 1[Y \leq y] \). In this case, the proof above does not work anymore as one cannot fully integrate out \( V \). One can obtain partial identification results following Imbens & Newey (2009), but the bounds can become quite large. With the first result from this section one can find control variables for which the large support assumption is satisfied, or at least close to being satisfied, such that one can integrate out a large portion of \( V \) and hence obtain tight bounds for partial identification results.

General control variable approach for univariate \( X \) The key for a generalization of the control variable approach for univariate \( X \) is to replace the map \( F_{X\mid Z}(X, Z) \) by a more general \( g^{-1}(X, Z) \), which allows for more general \( V \). The important requirement on \( g^{-1} \) is that it can be written as \( g^{-1}(X, Z) = f(h^{-1}(X, Z)) \) for some measurable function \( f \) whose inverse is also measurable, which must not depend on \( z \). Then \( f \) in the more general setting replaces the map \( F_U \), so that (5) takes the form

\[ V = g^{-1}(X, Z) = f(h^{-1}(X, Z)) = f(U), \quad (7) \]

and since \( f \) is measurable with measurable inverse and does not depend on \( z \), the \( \sigma \)-algebras of \( V \) and \( U \) must coincide.

It turns out that in the univariate case (7) is satisfied by a wide variety of distributions without stronger assumptions:

\[ 1[\cdot] \] is the standard indicator function, which is 1 if the statement in the brackets is true and 0 otherwise.
Multivariate Normal distribution, the large support assumption would automatically be satisfied for instance.

\(\text{is no longer the case that the half-open rectangles (} -\infty, u \text{)} \quad \text{probabilities if} \quad u \neq g(\Phi) \text{ and take} \quad h \quad \text{continuous functions to construct the control variable. Theoretically, however, for the control in higher dimensions. We can hence no longer rely on the monotonicity of strictly increasing and continuous and let Assumption 1 hold. If} \quad F_{X|Z=z} \quad \text{is univariate and continuous and} \quad h(z,u) \quad \text{is strictly increasing and continuous in} \quad U, \quad \text{then any univariate random variable} \quad V \quad \text{independent of} \quad Z \quad \text{with absolutely continuous and strictly increasing distribution} \quad F_{V} \quad \text{is a control variable if we take} \quad g^{-1}(x, z) \quad \text{to be} \quad F_{V}^{-1}(F_{X|Z=z}(x))\). 

This is a direct generalization of the control variable approach in Imbens & Newey (2009) as it allows for more control variables without making stronger assumptions. The key for the proof is the fact that the monotone rearrangement \(g^{-1}(Z,V)\) as well as the strictly increasing and continuous function \(h(Z,U)\) map, for all \(z \in Z\), each interval \((−∞,x]\) in \(\mathcal{X}\) to some interval \((−∞,h^{-1}(x,z])\) and \((−∞,g^{-1}(x,z)]\) in \(\mathcal{U}\) and \(\mathcal{V}\), respectively, as in laid out above in equation (4).

Theorem 1 is an important generalization of the control variable approach in practice since it allows the researcher to choose among a wide variety of distributions of \(V\) while \(X\) follows a bivariate Normal distribution, the large support assumption would automatically be satisfied for instance. We illustrate this below, where we apply our approach to the 1996 UK Expenditure survey.

**Theorem 1** (Control variable—univariate case). Let \(P_U\) be univariate and absolutely continuous and strictly increasing and let Assumption 2 hold. If \(F_{X|Z=z}\) is univariate and continuous and \(h(z,u)\) is strictly increasing and continuous in \(U\), then any univariate random variable \(V\) independent of \(Z\) with absolutely continuous and strictly increasing distribution \(F_{V}\) is a control variable if we take \(g^{-1}(x, z)\) to be \(F_{V}^{-1}(F_{X|Z=z}(x))\). 

\[P_U((−∞, u]) = P_{X|Z=z}((−∞, x]) = P_{X|Z=z'}((−∞, x']) = P_V((−∞, v]) \tag{8}\]

and the strict monotonicity of \(P_V\) and \(P_U\) implies that for each probability \(F_V(v)\) there exists a unique interval \((−∞, v)\) such that \(F_V(v) = P_V((−∞, v))\). The formal proof is in the appendix.

**Theorem 1** is an important generalization of the control variable approach in practice since it allows the researcher to choose among a wide variety of distributions of \(F_V\). Having the Freedom of choosing between different distributions makes it easier to satisfy the strong large support requirement \(\mathcal{V}_x = \mathcal{V}\) for all \(x \in \mathcal{X}\). For example, one can choose \(V\) to have a Normal distribution \(\Phi\) and take \(g^{-1}\) to be the map \(\Phi^{-1}(F_{X|Z=z}(x,z))\). Then in the event that \((V,X)\) follows a bivariate Normal distribution, the large support assumption would automatically be satisfied for instance.

We illustrate this below, where we apply our approach to the 1996 UK Expenditure survey.

**Multivariate X** A generalization to the multivariate setting is much more demanding since it is no longer the case that the half-open rectangles \((−∞, u]\) and \((−∞, u']\) necessarily have different probabilities if \(u \neq u'\) for strictly increasing \(F_U\), simply because there is no natural complete order in higher dimensions. We can hence no longer rely on the monotonicity of strictly increasing and continuous functions to construct the control variable. Theoretically, however, for the control variable approach to work, one still needs to assume that \(h(Z,u)\) and \(g(Z,v)\) are invertible and preserve measure in the sense that

\[P_U(h^{-1}(E, z)) = P_{X|Z=z}(E) \quad \text{and} \quad P_V(g^{-1}(E, z)) = P_{X|Z=z}(E) \tag{9}\]

\[^{8}F_{V}^{-1}(q) := \inf\{v \in \mathbb{R} : F_{V}(v) > q\} \text{ is the quantile function.} \quad F_{V}^{-1}(F_{X|Z=z}(x)) \text{ is called the monotone rearrangement of} \quad F_{X|Z=z} \quad \text{to} \quad F_{V}.\]
for all Borel sets $E \in \mathcal{B}_{\mathbb{R}^d}$. Invertible maps which preserve measure abound, even in higher dimensions. The problem in this setting is to find measure preserving maps which also fix an order in the sense that $f(E) = g^{-1}(h^{-1}(E, z), z) = g^{-1}(h^{-1}(E, z'), z')$ is always the same map for all $z, z' \in Z$ as in the univariate case.

It turns out that standard and natural functional form assumptions on $h$ cannot guarantee this strong requirement (7) in a truly multivariate set-up since $h$ usually depends on the form of the distributions $F_U, F_X|Z=z$, and $F_V$ in contrast to the univariate case; the reason for this, again, is the lack of a complete order in higher dimensions. For instance, under the assumption that $h(z, u)$ is the gradient of some convex function $\varphi$, i.e. $h(z, u) := \nabla \varphi_z(u)$, the arguably most natural generalization of a strictly increasing and continuous function $h_{10}$ the relation (7) does not hold unless $U = V$ in which case the control variable approach is rendered obsolete $^{11}$.

Therefore, the only way to generalize the control variable approach to higher dimensions is by either making strong functional form assumptions on $h$ as argued in H"oderlein, Holzmann, Kasy & Meister (2015), or making drastically weaker assumptions on $h$ under a normalization of $f$. We go the latter route. The key here is the famous result from Brenier (1991) about multivariate rearrangements of vector-valued functions (also see Chapter 3 in Villani 2003).

In probability terms Brenier’s result can be stated as follows. Pick two random variables $U$ and $V$ with absolutely continuous probability measures $P_U$ and $P_V$ on $\mathbb{R}^d$ and some function $f : U \subseteq \mathbb{R}^d \rightarrow V \subseteq \mathbb{R}^d$ which is square integrable with respect to $P_U$. In addition, assume there is another random variable $X$ with absolutely continuous probability measure $P_X$ on $\mathbb{R}^d$. Then $f$ can be uniquely factored as $f = \nabla \varphi \circ s$, where $\nabla \varphi$ is the unique gradient of a convex function mapping $X$ to $V$ and $s$ is a function mapping $U$ to $X$ and preserving measure in the sense of (9), i.e. $P_U(s^{-1}(E)) = P_X(E)$ for every Borel set $E \in \mathcal{B}_{\mathbb{R}^d}$. Consider Figure 4.

$$\begin{align*}
P_V \\
\nabla \varphi \\
P_X &\downarrow f \\
s &\downarrow \\
P_U
\end{align*}$$

Figure 4: Schematic presentation of Brenier’s rearrangement theorem

This relation holds for any absolutely continuous $X$; the only things that change for different $W_z$

---

$^9$Of course, a simple multivariate extension is to consider $U = (U_1, \ldots, U_d)'$ and $V = (V_1, \ldots, V_d)'$ as vectors of univariate elements $U_i$ and $V_i$ and simply stack $d$ univariate strictly increasing and continuous $h_i$ and $g_i$ onto one another to generate multivariate maps $h = (h_1, \ldots, h_d)'$ and $g = (g_1, \ldots, g_d)'$ from $U$ to $V$. In this case, everything reduces to the standard univariate case. This, however, can hardly count as a truly multivariate map, as there are no relations between the different elements of the vectors.

$^{10}$Note that a strictly increasing and continuous function is the derivative of its strictly convex antiderivative, see for instance Gunsilius & Schennach (2017).

$^{11}$The proof of this statement is rather technical and is available upon request.
are the unique maps $\nabla \varphi_z : X_z \to V$ and $s_z : U \to X_z$. Comparing Figures 3 and 4 provides us with the assumptions for a multivariate generalization of the control variable approach when setting $g(z, \cdot) = \nabla \varphi$, and $h(z, \cdot) = s_z$. The idea is a shift in perspective by not fixing $h$ but rather $f$. Recall that in practice $h$, $f$, and $U$ are unobservable while $X_z$ is and $V$ is constructed. Formally, one needs to consider the equivalence class of all maps between fixed $X$ and the map $g$ setting us with the assumptions for a multivariate generalization of the control variable approach when conditioning on $V$ is the same as conditioning on $U$ is constructed.

**Theorem 2** (Multivariate case). In model 1 let $P_U$ and $P_{X|Z=z}$ be absolutely continuous with finite second order moments for all $z \in Z$ and let Assumption 1 hold. Let $h(z, \cdot) : U \subseteq \mathbb{R}^d \to X_z \subseteq \mathbb{R}^d$ only be determined up to the fact that it preserves measure between $P_U$ and $P_{X|Z=z}$ in the sense of 9 and is invertible with $h^{-1}(x,z)$ preserving measure between $F_{X|Z=z}$ and $F_U$. Then every $V$ with absolutely continuous probability measure $P_V$ in $\mathbb{R}^d$ independent of $Z$ can be a control variable if we pick $g^{-1}(x,z)$ to be the unique gradient of a convex function preserving measure between $F_{X|Z=z}$ and $F_V$ and require $h$ to be such that $f = g^{-1}(h(u,z), z)$ for all $z$ for some fixed $f : U \to V$.

Theorem 2 is the natural generalization of Theorem 1 as it replaces the monotone rearrangement $g := F_V^{-1}(F_{X|Z=z}(x))$ from the univariate case by the gradient of a convex function $\nabla \varphi$ in the multivariate case.

The necessarily weak functional form requirement on $h$ might be unsatisfactory in some cases. In fact, instead of making stronger functional form assumptions on $h$, we make this assumption on $g$ to generate $V$ while being oblivious to the actual functional form of $h$ besides the fact that it is measure preserving. Slightly stronger functional form assumptions on $h$, like assuming that it too is the gradient of some convex function, do not lead to a reasonable control variable approach as then $f$ necessarily changes with $z$ as mentioned above. So in practice, if one wants to include several endogenous variables in a fully nonseparable triangular model, one has two choices: either one makes explicit functional form assumptions on $h$ like linearity or one generates an absolutely continuous random variable $V$ and is content with assuming that there exist $f$ and unique corresponding measure preserving $h(z, \cdot)$ which make $V$ a control variable.

**Summary of the control variable approach** The control variable approach is introduced because one does not want to make a distributional assumption on the unobservable $U$ for identification of average or marginal effects of the second stage. For a univariate and absolutely continuous $U$, we show that any absolutely continuous univariate random variable $V$ can serve as a control variable if we maintain the rather strong functional form assumption of Imbens & Newey (2009) that the production function $h$ in $X = h(z, U)$ is strictly increasing and continuous for all $z$. That is, no matter the distribution of $U$, we can choose an absolutely continuous $V$ such that conditioning on $V$ is the same as conditioning on $U$. The key is the strict monotonicity of $h$ and the map $g_z(x) := F_V^{-1}(F_{X|Z=z}(x))$ in conjunction with the total order in one dimension.
In a multivariate setting, due to the lack of a total order, the control variable approach requires stronger assumptions. The standard results in the literature usually make parametric functional form assumptions on \( h \), but we show that if one is willing to only assume that \( h \) is measure preserving, then one can directly generalize the univariate result, by replacing \( g_z := \nabla \varphi_z(x) \) as the gradient of a convex function, called the *Brenier map*. Brenier’s rearrangement theorem then not only guarantees that \( g_z \) is the unique map with this functional form between \( F_{X|Z=z} \) and \( F_V \), but also that for given \( f : U \rightarrow V \) and \( F_{X|Z=z} \), there is a unique pair \((g_z, h(z, \cdot))\). The difference from the univariate to the multivariate setting is hence that in the latter one does not know anything about the first stage production function besides the fact that it is measure preserving.

One can also interpret Theorem 2 informally, in which case it says that no harm is done by assuming that \( h \) is the Brenier map between the unobservable \( U \) and \( X \) since for every \( U \) and \( f \) there exists a Brenier map, such that \( V \) is a control variable. In the univariate case, one needs to assume that \( h \) is strictly continuous and increasing to make it work. The focus for identification in many settings is the production function \( m \) of the second stage, so that being oblivious about the true first stage relationship might be a reasonable thing to do. In section 4 we apply our two results to the 1996 UK Family Expenditure Survey to estimate quantile structural functions of Engel curves, analogous to [Imbens & Newey (2009)](https://doi.org/10.1093/ecta/88.1.34) and the semiparametric result in [Blundell, Chen & Kristensen (2007)](https://doi.org/10.1093/acprof:oso/9780198566121.001.0001).

The control variable approach is helpful in many settings, but very often researchers actually want to make sure that \( m \) itself is point identified. This is important in demand estimation, for identification of utility functions, for example. The following result generalizes the seminal point-identification result of [Torgovitsky (2015)](https://doi.org/10.1093/qje/90.3.657) to allow for a true multivariate first- and second stage.

### 3.2 Point-identification using discrete instruments

This subsection contains the main result of the article, where we generalize the seminal point-identification approaches from [Torgovitsky (2015)](https://doi.org/10.1093/qje/90.3.657) and [d’Haultfœuille & Février (2015)](https://doi.org/10.1093/qje/90.3.657) based on our general framework. As in the previous section, we rely on the concept of measure preservation to arrive at the general identification result. As mentioned in the introduction of section 3, the class of measure preserving isomorphisms between two probability measures is rather large, so that one needs to make further functional form assumptions in order to point-identify \( m \). We thus require that \( m \) be a *determinable* measure preserving isomorphism between \( F_{Y|X=x} \) and \( F_\varepsilon \), continuous in \( X \).

**Definition 3.** We call the measurable production function \( m : \mathcal{E} \rightarrow \mathcal{Y} \) **determinable** if the class of functional form assumptions on \( m \) intersected with the class of measure preserving isomorphisms between \( P_\varepsilon \) and \( P_{Y|X=x} \) contains a unique element for all \( x \in \mathcal{X} \).

Note that requiring \( m(x, \varepsilon) \) to be strictly continuous and increasing in \( \varepsilon \) is a functional form restriction which makes it determinable, if we in addition rule out the existence of strictly increasing
and continuous transformations \( g \circ m \) of \( m \), because for given \( F_\varepsilon \) \( m \) and \( g \circ m \) are observationally equivalent (Matzkin 2003, Lemma 1).

Strictly increasing and continuous functions are, however, only one possible form restriction, as mentioned in the footnote on page 7. Consider the following example in the univariate case.

**Example 1** (A matching example on the real line with concave cost functions). McCann (1999) introduces a matching model on the real line \( \mathbb{R} \) between a distribution \( P_\varepsilon \) of suppliers (e.g. coal mines) of some product and a distribution \( P_Y \) of demanders (e.g. factories) of this product. The problem is this model consists of minimizing the transport cost between coal mines and factories. The cost of transportation is modeled as \( c(y - e), \ y \in \mathcal{Y}, \ e \in \mathcal{E} \) for strictly concave \( c \). McCann argues that a concave cost function of the transport distance provides a reasonable model for applications in which shipping occurs along a single route, because the resulting shipping routes display economies of scale. He goes on to show that solving this optimal transport problem under the specified concave cost function has a unique measure preserving isomorphism as a solution \( Y = m(\varepsilon) \), which is neither continuous nor monotone. Assuming that \( m \) is the solution of the above optimal matching for concave costs is a functional form restriction which makes \( m \) determinable.

Since \( m \) is neither continuous nor strictly increasing, it cannot be identified by current approaches in the literature, however. Allowing for determinable functions \( m \) instead of only strictly increasing and continuous is especially important in the multivariate case as there are many determinable maps in this case which are not strictly increasing, many stemming from the theory of optimal transport (Galichon 2016). The most relevant example of a multivariate determinable production function is the Hedonic model with general heterogeneity and endogenous characteristics for which we prove identification in section 4.

Monotonicity also leads to determinability in higher dimensions.

**Example 2** (Multivariate monotonicity). By Brenier’s theorem mentioned in the previous section, for absolutely continuous probability measures \( P_\varepsilon \) and \( P_{Y|X=x} \) with finite second order moments there exists a unique measure preserving isomorphism \( m(x, \cdot) \) transporting \( P_\varepsilon \) onto \( P_{Y|X=x} \) for all \( x \), which takes the form of a gradient of a convex function, i.e. \( m(x, \cdot) := \nabla \varphi_x(\cdot) \) for convex \( \varphi_x \).

McCann (1995) generalized Brenier’s theorem to show that this result holds even if the probability measures do not have finite second order moments. We have included this result in the appendix (Theorem 5) for the sake of completeness. Therefore, assuming that \( m(x, \cdot) \) is the gradient of a convex function for all \( x \in \mathcal{X} \) is a functional form restriction which makes \( m \) determinable.

Gradients of convex functions are the most natural generalization of increasing and continuous functions. In particular, if \( T = \nabla \varphi \) for some convex \( \varphi \), then it is monotone in the following sense (Villani 2003, p. 53):

\[
\langle \nabla \varphi(x) - \nabla \varphi(z), x - z \rangle \geq 0,
\]
where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{R}^d$. Here it is easy to see that if $x > z$ in the partial ordering induced by the positive cone on $\mathbb{R}^d$, then this definition implies that $\nabla \varphi(x) \geq \nabla \varphi(z)$ or that $\nabla \varphi(x)$ and $\nabla \varphi(z)$ are not comparable. This monotonicity property has been exploited recently by Carlier, Chernozhukov & Galichon (2016a) who use it to generalize the concept of univariate quantiles.

In general, functional form restrictions will most likely come from solving functional equations in economic theory, like a minimization problem for demand functions, for example. If these functional equations admit a unique invertible solution, then they induce a functional form restriction which make their solution determinable.

**Example 3 (Demand function).** Following Matzkin (2007a), we denote by $V(y, x, \varepsilon)$ the indirect utility function of a consumer over bundles of goods $y$, where $x$ are observable and $\varepsilon$ are unobservable characteristics. Then a demand function can be obtained by

$$d(p, I, x, \varepsilon) := \arg\min_y \{V(y, x, \varepsilon) : p \cdot y \leq I\},$$

for $p$ the price vector and $I$ the initial endowment. The standard assumption is then that $d$ is the unique solution and invertible in $\varepsilon$ (e.g. Berry & Haile 2014, p. 1757), which makes $d$ determinable.

Note that in order to identify a determinable $m$ in practice, one needs to make a normalization assumption, usually on the unobservable distribution $F_\varepsilon$, to guarantee that there is only a unique set of $(m, F_\varepsilon)$ which can generate the distribution $F_{Y|X=x}$ for all $x \in X$. This should be intuitively clear as one in principle needs to identify two things, the unobservable $F_\varepsilon$ as well as the corresponding production function $m$.

**Point-identification of the second stage in nonseparable triangular models** We can now lay out the assumptions for the main identification result. A convenient property of our approach is that one can use the assumptions from Torgovitsky (2015a) for the first stage $X = h(Z, U)$, i.e. requiring that $h$ can be written element-wise as $h(Z, U) = [h_1(Z, U_1), \ldots, h_k(Z, U_k)]'$ and require the univariate functions $h_i$ to be strictly increasing and continuous in each univariate $U_i$, which makes our approach a direct generalization. This, however, requires the strong assumption of a compact and rectangular support for $X|Z = z$ for all $z \in Z$ (Torgovitsky 2015b and d’Haultfoeille & Février 2015).

We therefore propose an alternative approach which allows for weaker assumptions on the support of $F_{X|Z=z}$ and $h$, but requires an additional normalization assumption on the first stage, which either directly or indirectly fixes the distribution of $U$. The first way to do this is by using the multivariate version of the control function approach introduced in the previous section. In this case, we do not have to make any assumptions on $U$, because we can construct an appropriate control variable $V$ which mimics the behavior of $U$. The cost in this case is to only know that
$h(z, \cdot)$ is measure preserving, but not being able to pin down more precise functional forms. In this case we would have to work with $V$ instead of $U$ throughout what follows.

An alternative and more straightforward way is to simply assume that the distribution of $U$ is known. With this assumption we can make more restrictive functional form assumptions on $h$, like requiring it to be the gradient of a convex function itself. Both approaches work equally well, in the control variable approach we do not make assumptions on $U$ and rather construct $V$ and $g$, where $g$ is the Brenier map (i.e. the gradient of a convex function), whereas in the more direct approach we assume $U$ is known and require $h$ to be the Brenier map between $U$ and $X$. Neither approach requires strong assumptions, and while the control variable approach is likely to be more natural, we go with the second approach and assume $U$ is known and $h$ is the Brenier map between $U$ and $X$, simply because we think it is easier for the reader to follow our argument. We therefore have the following assumptions.

**Assumption 3.** The production function $h(z, u)$ in first stage of model is the Brenier map between $U$ and $X$ for all $z \in Z$. Moreover, $U$ has an absolutely continuous distribution $F_U$.

The following assumption is the normalization assumption on $U$.

**Assumption 4.** There is a known $\bar{z} \in Z$ such that $h(\bar{z}, u) = u$ for all $u \in U$.

Assumption 4 is a standard assumption made in the literature and was first proposed in Matzkin (2003). It requires $h$ to be the identity mapping for a known $\bar{z}$ between $X$ and $U$ and hence fixes the distribution of $U$. This assumption has also been used in other areas, most notably the measurement error literature, where Hu & Schennach (2008) require a known functional which fixes the unobservables in a certain rotation. Assumption 4 allows us to dramatically relax the strong assumption of a rectangular and compact support on the observable $F_X|Z=z$:

**Assumption 5.** The distributions $F_{X|Z=z_i}$ are absolutely continuous with convex support $X_{z_i} \subseteq \mathbb{R}^k$ for all $z_i \in Z$ and are strictly increasing.

We allow for $Z$ to be a discrete and even binary instrument. In the following we focus on the binary case, i.e. $Z = z, z'$, because in the proof of the main result we assume $h$ to be the identity for one $z$ and the gradient of a convex function for the other $z'$. Also, for the identification of Hedonic models, two markets are sufficient by definition, and adding more markets would not change the result in any way.

**Assumption 6.** $Z$ is a valid instrument for $X$, i.e. (i) it generates exogenous variation in $X$ such that $F_{X|Z=z}(x) \neq F_{X|Z=z'}(x)$ for at least one $x \in X$ and (ii) is independent of $\varepsilon$ and $U$, denoted by $(\varepsilon, U) \perp \perp Z$.

---

12 It will be important to work with the Brenier map for our main identification result. In particular, the whole identification result rests on an apparently new insight into properties of the Brenier map between two multivariate probability distributions, which we prove in Lemma 1 in the appendix.

13 We have included the statement and the proof of the theorem for continuous $Z$ in the appendix, because it is not relevant for the identification of multi-market Hedonic models.
Note that we do not require the dimensionality of Z to be at least the dimensionality of X. In fact, and this is analogous to the result in Torgovitsky (2015b), Z can be lower- and even one-dimensional for multivariate X and the approach still works. This is important for the application to Hedonic models, because the idea is to consider each market to be the realization of a one-dimensional instrument. Intuitively, we can allow for lower-dimensional Z because we make use of the nonlinear and nonparametric structure, i.e. taking into account the information of all higher order and not just first-order moments; this will become apparent momentarily when we introduce Assumption (9).

The ultimate goal for identification is to point-identify the second stage production function m. Our identification result allows m to be multivariate.

**Assumption 7.** m(x, ε) is a determinable measure preserving isomorphism between Y and ε and is continuous in X for all x. Moreover, Pε is known.\(^{14}\)

It turns out, however, that assuming continuity of m(x, ε) in x is a rather strong assumption. In particular, we will not be able to guarantee this in our efforts to identify the multi-market Hedonic model in the next section. Therefore, we need to make a weaker assumption. This requires the definition of convergence in measure, see for instance Bogachev (2007a, Definition 2.2.2).

**Definition 4.** Suppose we are given a measure space (X, 𝒜X) with a probability measure P and a sequence of P-measurable functions \(\{f_n\}_{n \in \mathbb{N}}\). Then the sequence \(\{f_n\}_{n \in \mathbb{N}}\) is said to converge in measure to a P-measurable function f if for every c > 0 one has

\[
\lim_{n \to \infty} P(x : |f_n(x) - f(x)| \geq c) = 0.
\]

Based on this, we only require m(x, ε) to be continuous in x in the following weaker sense.

**Definition 5.** If for every sequence \(\{x_n\}_{n \in \mathbb{N}} \in X\) which converges to some x ∈ X the corresponding sequence m(x_n, ε) converges in measure to m(x, ε), we say that m is continuous in measure.

The weaker assumption we need to require for m is hence that it is continuous in measure based on Definition 5.

**Assumption 5’.** m(x, ε) is a determinable measure preserving isomorphism between Y and ε and is continuous in measure in x. Moreover, Pε is known.

We also have to make a large support assumption on ε. This is the same assumption that Torgovitsky (2015a) had to impose.

\(^{14}\)In the univariate case following Matzkin (2003), Pε is usually normalized to be the uniform distribution. Since we work with general and multivariate distributions, we only assume it is known. Alternatively, instead of assuming Pε is known, one could assume that m(\(\bar{x}, e\)) = e at some known \(\bar{x} \in X\) for all e ∈ E, i.e. assuming that it is the identity map for \(\bar{x}\), just as for the first stage. Then Pε is also fixed, because by measure preservation of m we have P_{Y|X=\bar{x}}(E) = P_ε(m^{-1}(\bar{x}, E)) = P_ε(E) for every Borel set E in \(\mathcal{B}_\mathbb{R}^d\).
**Assumption 8.** The support $E$ is convex and independent of $X = x$ and $Z = z$, i.e. $E_{x,z}$ coincides with $E$ for all $(x, z) \in X \times Z$.

Now, in order to identify the model with discrete or binary instruments, the main assumption needs to be made on the distribution functions $F_{X|Z=z_i}$, $i = 1, \ldots, k_z$. It is the analogous assumption to the ones made in Torgovitsky (2015), who requires all univariate distribution functions to intersect in at least one point. In the following we consider binary $Z$ with realizations $z$ and $z'$; furthermore, for each pair $z, z' \in Z$ we denote as $I(z, z') \subseteq X_z \cup X_{z'}$ the set where $F_{X|Z=z}$ and $F_{X|Z=z'}$ intersect. That is:

$$I(z, z') := \{ x \in X_z \cup X_{z'} : 0 < F_{X|Z=z}(x) = F_{X|Z=z'}(x) < 1 \}.$$

Moreover, for every point $x_0 \in X_z = X_{z'}$, we let $I_z(x_0)$ denote the isoquant or level set of $F_{X|Z=z}$ at $x_0$, which is the set of all $x \in X_z$ which have the same probability as $x_0$ under $F_{X|Z=z}$, formally:

$$I_z(x_0) := \{ x \in X_z : F_{X|Z=z}(x) = F_{X|Z=z}(x_0) \}.$$

Since we assumed that $F_{X|Z=z}$ is absolutely continuous with convex support and strictly increasing in a multivariate sense, we can regard the distribution functions $F_{X|Z=z}$ and $F_{X|Z=z'}$ as utility functions, in which case the isoquants $I_z(\cdot)$ and $I_{z'}(\cdot)$ can be interpreted as the indifference curves (or in higher dimensions: indifference manifolds) of $F_{X|Z=z}$ and $F_{X|Z=z'}$, respectively. This analogy is the key in proving the result as it allows us to work with the indifference curves instead of the distribution functions.

Now for stating the main assumption which gives us identification, we need to introduce the concept of transversal intersection of manifolds. The following definition is adapted from Milnor (1997).

**Definition 6.** Two submanifolds $N$ and $N'$ of an ambient manifold $M$ intersect transversally if for each $x \in N \cap N'$ their tangent spaces at $x$, denoted by $T_x N$ and $T_x N'$, together generate the tangent space $T_x M$ in the sense that $T_x N + T_x N' = T_x M$.

Transversal intersection of indifference curves of different utility functions is a standard assumption made in Economic Theory (Mas-Colell 1989) and is very weak since it is a generic property in the sense that basically all indifference curves between different preferences intersect transversally by a result from René Thom (see Ekeland, Heckman & Nesheim 2004 for a discussion of generic properties). We are now in the position to state the main assumption on the distributions $F_{X|Z=z}$.

**Assumption 9.** Let $Z$ be binary with $Z = \{ z, z' \}$. Then the following properties of $F_{X|Z=z}$ and $F_{X|Z=z'}$ hold:

1. $X_z = X_{z'}$.  

18
2. The epigraphs\footnote{The epigraph of a real valued function \( f : X \to \mathbb{R} \) for the level \( \alpha \in \mathbb{R} \) is defined by \( \text{epi}(f; \alpha) \equiv \{(x, \alpha) \in X \times \mathbb{R} : \alpha \geq f(x)\} \), see Aliprantis & Border (2006).} of \( F_{X|Z=z} \) and \( F_{X|Z=z'} \), denoted by \( \text{epi}(F_{X|Z=z}; \alpha) \) and \( \text{epi}(F_{X|Z=z'}; \alpha) \) for \( \alpha \in [0,1] \), are convex sets. The set of all points where the isoquants meet, \( \mathcal{I}(z, z') \), consists of at least one connected manifold \( \mathcal{M}(z, z') \). At all points \( x_0 \in \mathcal{I}(z, z') \) the isoquants either intersect transversally or coincide in a neighborhood \( \mathcal{N}(x_0) \) around \( x_0 \).

3. The manifold \( \mathcal{M}(z, z') \) is such that

(i) for each \( x \in \mathcal{X}_z = \mathcal{X}_{z'} \) there is an \( m \in \mathcal{M}(z, z') \) with \( 0 < F_{X|Z=z}(m) = F_{X|Z=z'}(m) < 1 \) which either dominates or is dominated by \( x \).

(ii) All points \( x \in \mathcal{X}_z \) with \( F_{X|Z=z}(x) = 0 \) lie on one side of the manifold and all points where \( F_{X|Z=z}(x) = 1 \) lie on the other, and analogously for \( F_{X|Z=z'}; \) “lying on one side of the manifold” means that there are no two points \( x_1, x_2 \in \mathcal{X}_z \) with \( F_{X|Z=z}(x_1) = F_{X|Z=z}(x_2) = 0 \) (respectively: \( F_{X|Z=z'}(x_1) = F_{X|Z=z'}(x_2) = 1 \)) such that the line \((1-t)x_1 + tx_2 \) for \( t \in [0,1] \) intersects \( \mathcal{M}(z, z') \).

Part 1 of Assumption 9 is restrictive as it requires that the supports of all conditional distribution functions coincide. This assumption is slightly stronger than the assumption in the univariate case of Torgovitsky (2015\textsuperscript{a}) as the distribution functions there only need to intersect but the supports need not coincide. Parts 2 and 3 of Assumption 9 appear to be high level, but are actually rather natural and weak.

To see that Assumption 9 is a reasonable assumption to make in practice, consider Figure 5 as an example, where we display the intersection of a bivariate \( t \) distribution \( F_{X|Z=z} \) with density function

\[
f_{X|Z=z'} = (v\pi)^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \frac{1}{\Gamma(\frac{1}{2}v)} \left( 1 + \frac{x^T \Sigma^{-1} x}{v} \right)^{-\frac{1}{2}v-1} \quad \text{for } v = 2 \text{ and } \Sigma = \begin{pmatrix} 2 & 0.8 \\ 0.8 & 0.5 \end{pmatrix}
\]

with \( \Gamma(z) = \int_{0}^{\infty} x^{z-1} \exp(-x) dx \)

and a bivariate Normal distribution \( F_{X|Z=z'} \) with density function

\[
f_{X|Z=z} = (2\pi)^{-\frac{1}{2}} |\Sigma'|^{-\frac{1}{2}} \exp(-\frac{1}{2}(x - \mu)^T (\Sigma')^{-1} (x - \mu)) \quad \text{for } \mu = (0, 0)' \text{ and } \Sigma' = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 4 \end{pmatrix}.
\]

In this case, the set where all isoquants intersect, \( \mathcal{I}(z, z') \), consist of two separate manifolds. The one on the bottom does not satisfy part 4 of Assumption 6 but the one on top, \( \mathcal{M}(z, z') \), does. This is all we need since we only need one manifold to satisfy Assumption 9. To check part 3 of Assumption 9 we need to look at the contour plot. Figure 6 shows that the respective isoquants either intersect transversally (in the interior of the graph) or converge towards one another (at the boundaries) so that they coincide there. Assumption 9 is hence satisfied which...
Figure 5: Example of a bivariate Normal distribution and a bivariate t distribution

would guarantee point-identification of \( m \) in a nonseparable triangular model where \( F_{X|Z=z} \) and \( F_{X|Z=z'} \) are those two distributions. We did not explicitly engineer this example, it is simply the most straightforward and most relevant application we can imagine, because the multivariate t distribution is very similar to a Normal distribution, but has thicker tails and both are pervasive. We obtain similar pictures if we consider the intersection of two Normal distributions, so that identification also holds there. This makes us confident that Assumption 9 is satisfied in many important practical applications, not just in two but also higher dimensions. That said, there are certainly cases which do not satisfy Assumption 9. In particular, we cannot allow for the fact that one distribution first-order stochastically dominates the other distribution in a multivariate sense, but this requirement is the same as in Torgovitsky (2015) or d’Haultfœuille & Février (2015), only in the multivariate case.

In the univariate case, i.e. when \( X, U \in \mathbb{R} \), the manifold \( M(z, z') \) reduces to a point in the special case where \( \mathcal{X}_z \) and \( \mathcal{X}'_{z'} \) coincide. In this case we can simply use Torgovitsky’s assumption on the first stage which makes our approach a direct generalization of Torgovitsky (2015).

To get a better idea of when Assumption 9 holds in general, consider Figure 7 and suppose that the sheet of paper represents \( \mathbb{R}^2 \) with the standard partial order \( x \geq y \) if and only if \( x_1 \geq y_1 \) and \( x_2 \geq y_2 \). Depicted there are three different scenarios for the supports \( \mathcal{X}_z \) and \( \mathcal{X}'_{z'} \) as well

\(^{16}\)We want to mention in this respect that even though the form of \( I(z, z') \) depends on \( \Sigma \) and \( \Sigma' \), identification holds for basically all combinations of \( \Sigma_i \) and \( \Sigma_j \) and degrees of Freedom \( v \); that is, in all cases there is a manifold \( M(z, z') \subseteq I(z, z') \) with the required properties. Note that we have chosen an example where the variance of \( F_{X|Z=z} \) does not exist since \( v = 2 \). Our approach still works in this case since the Brenier map exists (see Theorem 5 in the appendix). Also, if we enlarge the grid—which in principle we would need to do since both supports are infinite—the manifold \( M(z, z') \) becomes more pronounced in the sense that it satisfies the requirements more easily. This should intuitively make sense as the \( t \) distribution has fatter tails than the multivariate Normal.
as the manifold $\mathcal{M}(z, z')$ in these supports. These pictures are schematic versions of Figure 6 in that they only depict the supports $\mathcal{X}_z$ and $\mathcal{X}_{z'}$ and the respective manifolds $\mathcal{M}(z, z')$, but not the contour plots. Consider the picture on the left first. Any one of those three depicted manifolds in this example satisfies Assumption 9 as all three are of dimension 1 and connected; moreover, for each point $x \in \mathcal{X}_z = \mathcal{X}_{z'}$, there is a point in every manifold which either dominates or is dominated by $x$. Also all points in the support with $F_{X|Z=z}(x) = 0$ lie on the same side of the manifolds, and analogously for $F_{X|Z=z'}$. The example in the center violates Assumption 9 since $x_1$ with $F_{X|Z=z}(x_1) = 0$ and $x_2$ with $F_{X|Z=z}(x_2) = 0$ lie on opposite sides of $\mathcal{M}(z, z')$. Finally, the example on the right violates Assumption 9 even though there are two connected manifolds which together are such that each point $x$ either dominates or is dominated by some $m$ in one of the manifolds. The problem here is that there is not one manifold alone for which this holds. In fact, there is no point in the top left manifold which either dominates or is dominated by $x_3$, i.e. lies to the north-east or south-west; similarly, there is no point on the bottom right manifold which either dominates or is dominated by $x_3$. In addition, both manifolds violate the requirement that all points with $F_{X|Z=z}(x) = 0$ and $F_{X|Z=z'}(x) = 0$ lie on one side.

Assumption 9 is weaker than the currently existing assumptions allowing for a multivariate $X$. In fact, we only require the supports of $F_{X|Z=z_i}$ to be convex, but they can be unbounded. Also, we only require one manifold $\mathcal{M}(z, z')$ for binary $Z$. In contrast, Torgovitsky (2015b) and d’Haultfoeuille & Février (2015) require the supports of $F_{X|Z=z_i}$ to be rectangular and compact,

\footnote{Recall that a point $x$ dominates $x'$, i.e. $x > x'$, if it lies to the “north-east” of $x'$.}

\footnote{Note that $F_{X|Z=z}(x_1) = 0 = F_{X|Z=z'}(x_2)$, because the rectangles $(-\infty, x_1]$ and $(-\infty, x_2]$ do not intersect the support, so that $F_{X|Z=z}(x_1) = P_{X|Z=z}((-\infty, x_1]) = 0 = P_{X|Z=z'}((-\infty, x_2]) = F_{X|Z=z'}(x_2)$.}

Figure 6: Contour plot corresponding to Figure 5
Figure 7: Examples (not) satisfying Assumption 9

A much stronger condition. Moreover, note that Torgovitsky (2015b) requires that for every endogenous variable $X_i$, $F_{X_i|Z=z}$ and $F_{X_i|Z=z'}$ have to intersect in at least one point, and this for every $i = 1, \ldots, d_x$. So in a $k$-variate case, this would actually require $k$-linear curves, each orthogonal to one of the $k$ dimensions, instead of one general curve from Assumption 9.

All Assumptions are rather weak and can even be checked by estimating the respective distribution functions $\hat{F}_{X|Z=z}$ and $\hat{F}_{X|Z=z'}$ and examining whether their intersection satisfy Assumption 9. Note, however, that, analogous to Torgovitsky (2015b), this support assumption excludes linear relationships like $X = \beta Z + U$ for $Z = \{0,1\}$, because in those relationships the two conditional distribution functions would simply be a shift of each other and would not intersect.

Still, by our above reasoning and the fact that the assumption is satisfied by Normal distributions, we are convinced that Assumption 9 and therefore the following theorem hold in many if not most practical settings.

**Theorem 3.** Let Assumptions 3–9 hold and let $Z$ be discrete with at least two points in its support. Then $m(x, \varepsilon)$ is identified for almost every $x \in X$ and all $\varepsilon \in E$ in model (1). The identified set

$$I := \{m \in H(Y_{xz}, \varepsilon_{xz}) : (m^{-1}(X,Y),U) \perp \!\!\!\perp Z\}$$

hence contains an $X$-almost everywhere unique element. Here, $H(Y_{xz}, \varepsilon_{xz})$ denotes the set of all measure preserving isomorphisms between $P_{Y|X,Z}$ and $P_{\varepsilon|X,Z}$ satisfying Assumption 7. If Assumption 7 holds in place of 7, the analogous result holds, but $m$ is then only identified for almost every $\varepsilon \in E$ instead of all $\varepsilon$.

The importance of Assumption 5 is that it basically does not weaken the result (identification for almost every $\varepsilon$ compared to identification for every $\varepsilon$) compared to Assumption 7 while being a rather substantial weakening of Assumption 7. In particular, we can guarantee continuity in probability of $m$ but not full continuity in the next section for identification of Hedonic models.
Intuition and outline of the formal proof  We have relegated the proof of this result to the appendix, but let us give an outline of the idea. Intuitively, the problem of identification results from the fact that \(m\) is the map between \(P_\varepsilon\) and \(P_{Y|X}\) for exogenous \(X\). If \(X\) were actually exogenous, we would not need a first stage relationship, because in this case the observable distribution \(F_{Y|X}\) is exactly the distribution corresponding to \(m\) and we could simply use the observable distribution and a normalization of \(F_\varepsilon\) to identify \(m\). This is the underlying idea for identification of single market Hedonic models with exogenous characteristics.

Since \(X\) is endogenous, however, the observable distribution \(F_{Y|X}\) is not the right distribution for identifying \(m\). We therefore need an instrument \(Z\) which has a nonzero influence on \(X\) and is independent of \(\varepsilon\). Then for a binary (or discrete) \(Z\) with values \(z\) and \(z'\) and the first stage relationships \(X = h(z, U)\) and \(X = h(z', U)\), we can use \(U\) with known distribution \(F_U\) as a control variable in the sense of Imbens & Newey (2009), only in a multivariate setting and not requiring \(U\) to have a uniform distribution. In fact, note that by the assumption that \(h\) is the Brenier map between \(X\) and \(U\) for \(z\) and \(z'\) and the identity map between \(X\) and \(U\) for \(z\) and the fact that both \(P_{X|Z}\) and \(P_U\) are absolutely continuous, \(h\) establishes a bijective relation between \(X\) and \(U\) for \(z\) as well as \(z'\). That is, for each \(u \in U\) there are two \(x, x' \in \mathcal{X}\), possibly coinciding, corresponding to it: \(u = h^{-1}(x, z)\) and \(u = h^{-1}(x', z')\). If we fix \(Z = z\), then the relation is bijective. In the other direction, for every \(x \in \mathcal{X}\) there are two \(u, u' \in U\): \(x = h(z, u)\) and \(x = h(z', u')\). Then the crucial ingredient is the measure preservation of \(h\). In fact, for every Borel set \(E_u \in \mathcal{B}_{\mathbb{R}^d}\) we have \(P_U(E_u) = P_{X|Z=z}(h^{-1}(E_u, z))\) and similarly for \(z'\), so that the distributions do not change if we condition on \(U\) or \(X\).

Therefore, for fixed \(Z\), conditioning on \(U\) is the same as conditioning on \(X\). Now the crucial assumption guaranteeing that \(Z\) is independent of \(\varepsilon\) and \(U\) allows us to use the fact that for each \(u\) there are two \(x, x'\), depending on which realization of \(Z\) we use for the map \(h\). Then the idea—for all \(x \in \mathcal{X}\)—is to construct a sequence from \(x\) to \(u\) via \(h^{-1}(\cdot, z)\), and then change to \(x'\) via \(x' = h(z', u)\), then change to \(u'\), and so forth. The key here is that for this sequence starting with any \(x\) the distributions \(F_{Y|X}\) and \(F_\varepsilon\) of the second stage do not change because of the independence of \(Z\) and \(\varepsilon\). So this sequence induces an exogenous change in \(X\) by changing \(Z\) which does not affect the distribution of \(F_\varepsilon\). By assumption 9 this sequence must converge and cannot go on forever, because at some point it must be that \(x = x'\). This holds for every starting point \(x\), so that we can in principle identify \(m\) by exogenously varying \(Z\).

In practice, we do not observe \(F_{Y|X}\) for exogenous \(X\), even using the instrument \(Z\). Theorem 3 hence only shows that \(m\) is identifiable, and gives us the identification set, but not a constructive way to obtain \(m\). This reasoning is perfectly analogous to and hence a generalization of the result in Torgovitsky (2015a) and also d’Haultfœuille & Février (2015). Note again that the dimension of \(Z\) can be smaller and even one-dimensional for this, as long as \(Z\) is a valid instrument for each variable in the vector \(X\). This works since we require nonlinearities of \(m\) by way of Assumption 9 and hence implicitly take into account the information of all higher order moments as mentioned.

To be slightly more formal: the basic idea is to prove uniqueness of \(m\). So it is natural to
assume that there are $m$ and $m^*$ as well as corresponding $\varepsilon$ and $\varepsilon^*$ satisfying the assumptions and $Y = m(X, \varepsilon)$ as well as $Y = m^*(X, \varepsilon^*)$. Identifiability can then be proved if $m = m^*$. This is done by showing that the isomorphism

$$q(x, z, \cdot) = q(x, \cdot) = m^{-1}(x, m^*(x, \cdot))$$

is actually the identity, i.e. that $q(x, e) = e$ for all $x \in \mathcal{X}_z \cup \mathcal{X}_{z'}$ and $e \in \mathcal{E}$, which would imply $m = m^*$ for every $x$. To show that $q(x, e)$ is the identity with respect to $\varepsilon$, it is actually sufficient to show that it is not a function of $x$ by the fact that $F_\varepsilon$ is known and $m$ is determinable. In fact, as $m(x, e)$ is determinable between $F_\varepsilon$ and $F_{Y|X=x}$ for each $x \in \mathcal{X}$, there can be no other $m(x, \cdot)$ of this functional form by definition for each $x$. Therefore, if $q(x, e)$ is only a function of $e$, say $f(e)$, this means that the functional form of $m^{-1} \circ m^*$ does not change with $x$ so that both $m$ and $m^*$ have the same functional form. But since $m$ is determinable, it must be that $m = m^*$. This is the same reasoning as in [Torgovitsky (2015a)], only put in a more general framework. In order to achieve this, Assumption 6 is crucial, as it guarantees that $P_{\varepsilon|X,Z} = P_{\varepsilon|X}$ and analogously for $\varepsilon^*$.

### Figure 8: Underlying isomorphism structure

Therefore, varying $Z$ does not affect the distributions of the unobservables, but does affect $X$, which means that one can get at the exogenous effect of $X$ on $Y$. This is the same reasoning as in the case where $Z$ is absolutely continuous. In the binary case one can use a general sequencing argument as mentioned above. Let us be more specific about this sequencing argument now.

Recall that in the univariate case [Torgovitsky (2015a)] uses the measure preserving isomorphism $(x, z) \mapsto (F_{X|Z=z}(x), z)$ to condition $\varepsilon$ on $F_{X|Z}$ instead of $X, Z$ and then applies the monotone rearrangement $T(x) = F_{X|Z=z}^{-1}(F_{X|Z=z}(x))$ as a map between $F_{X|Z=z}$ and $F_{X|Z=z'}$; this ensures that for every point $x_0 \in \mathcal{X}_z \cup \mathcal{X}_{z'}$, $F_{\varepsilon|X=x_0, Z=z} = F_{\varepsilon|X=T_{x_0}, Z=z}$ and analogously for $\varepsilon^*$, so that $q$ is the same for all iterations $T^n$. He then shows that in one dimension this iteration converges to a fixed point and can hence show that $q$ is constant for all starting points $x_0$, which by the assumed normalization implies that $m = m^*$.

Now, there are mainly two reasons for why this simple reasoning does not work in a higher dimensional setting. First, the map $(x, z) \mapsto (F_{X|Z=z}(x), z)$ is only invertible in the one-dimensional case, so this simple argument does not work. Our solution for this is Assumption 4. With this
we can write \( U = h^{-1}(X,Z) \) since both \( P_{X|Z} \) and \( P_U \) are absolutely continuous, so that \( h \) is invertible, which gives

\[
P_{\epsilon|X=x,Z=z} = P_{\epsilon|U=h^{-1}(x,z),Z=z} = P_{\epsilon|U=h^{-1}(x,z)}. \tag{10}
\]

The second equality in (10) follows from Assumption 9. The first equality follows from the following reasoning: the map \( \phi : (X,Z) \mapsto (h^{-1}(X,Z), Z) \) is a measure-preserving isomorphism since \( z \mapsto z \) is a measure preserving isomorphism and \( x \mapsto h^{-1}(x,z) \) is a measure preserving isomorphism for all \( z \), so that for every rectangle \( E_x \times E_z \equiv (-\infty, x] \times (-\infty, z] \in \mathcal{B}_{\mathbb{R}^{k+m}} \)

\[
P_{X,Z}(E_x \times E_z) = P_{U,Z}(\phi^{-1}(E_x \times E_z)) \equiv P_{U,Z}(h^{-1}(E_x, z) \times E_z).
\]

Analogously, the map \( (\epsilon, x, z) \mapsto (\epsilon, h^{-1}(x,z), z) \) is a measure preserving isomorphism for the same reasoning so that for every rectangle \( E_{\epsilon} \times E_x \times E_z \equiv (-\infty, \epsilon] \times (-\infty, x] \times (-\infty, z] \in \mathcal{B}_{\mathbb{R}^{d+k+m}} \)

\[
P_{\epsilon,X,Z}(E_{\epsilon} \times E_x \times E_z) = P_{\epsilon,U,Z}(\phi^{-1}(E_{\epsilon} \times E_x \times E_z)) = P_{\epsilon,U,Z}(E_{\epsilon} \times h^{-1}(E_x, z) \times E_z).
\]

Thus

\[
P_{\epsilon|X,Z}(E_{\epsilon}) = \frac{P_{\epsilon,X,Z}(E_{\epsilon} \times E_x \times E_z)}{P_{X,Z}(E_x \times E_z)} = \frac{P_{\epsilon,U,Z}(E_{\epsilon} \times h^{-1}(E_x, z) \times E_z)}{P_{U,Z}(h^{-1}(E_x, z) \times E_z)} = P_{\epsilon|U,Z}(E_{\epsilon}).
\]

The last thing to notice is that conditioning on measure zero events does not cause issues, because \((X,Z) \mapsto (h^{-1}(X,Z), Z)\) is measurable with measurable inverse by definition of a measure preserving isomorphism, so that their \( \sigma \)-algebras coincide, i.e. \( \sigma(U,Z) = \sigma(X,Z) \). We give another formal proof of this fact in Lemma 4 in the appendix, using disintegrations

Second, we need to use a general sequencing argument which is more intricate in higher dimensions. By Assumption 9, the distribution of \( U \) is fixed to be \( F_U = F_{X|Z=z} \) for one \( z \in \mathcal{Z} \) and we require the map \( h^{-1}(x,z) \) to be the identity and the other map \( h^{-1}(x,z') \) to be the Brenier map between \( F_U \) and \( F_{X|Z=z'} \). They key step here then is a new result for the dynamics of the Brenier map which we prove in the appendix. Intuitively, we use Assumption 9 and the strict monotonicity of the \( F_{X|Z} \) to show that the Brenier map \( T \) between \( F_{X|Z=z} \) and \( F_{X|Z=z'} \) never crosses \( \mathcal{M}(z, z') \) in the sense that for each \( x \in \mathcal{X}_z \cup \mathcal{X}_{z'} \) the curve \((1-t)x + tTx, t \in (0,1)\) never intersects \( \mathcal{M}(z, z') \). With this we can say that \( \mathcal{M}(z, z') \) is a fixed set of the iteration \( T^n x_0 \) for the Brenier map \( T \) between \( F_{X|Z=z} \) and \( F_{X|Z=z'} \). The key here is Assumption 9 which requires that the manifold \( \mathcal{M}(z, z') \) cuts the supports \( \mathcal{X}_z = \mathcal{X}_{z'} \) in such a way that an iteration of the

\footnote{Note that this conditioning is different from the approach in Kasy (2014). Kasy used the mapping \( \psi : (X,Z) \mapsto (X, h^{-1}(X,U)) \), where he defined the inverse of \( h \) is \textit{with respect to} \( X \) (!), which is not invertible since in his case it is a map from \( \mathbb{R}^d \) to \( \mathbb{R} \times U \), where \( U \) is the (in Kasy’s case possibly infinite dimensional) metric space containing \( U \). Therefore, the respective \( \sigma \)-algebras \( \sigma(X, Z) \) and \( \sigma(X, h^{-1}(X,U)) \) need not coincide. In our case, however, we use the measure preserving isomorphism \( \phi(X,Z) = (h^{-1}(X,Z), Z) \), which is measurable with measurable inverse, so that \( \sigma(X,Z) \) and \( \sigma(h^{-1}(X,Z), Z) \) coincide.}
Brenier map for every point converges to $\mathcal{M}(z, z')$. The proof of Theorem 3 contains the details.

Note in this respect that if we were to make a different functional form assumption on $h$ we would have to work out the dynamics of a different measure preserving isomorphism which in turn would lead to a different Assumption 9. The Brenier map is a very general and well-behaved map, however, and it is not likely that one will find a measure preserving map with better properties. Even more importantly, the assumption that $h$ is the Brenier map, i.e. the gradient of a convex function, is the most natural generalization of a strictly increasing and continuous $h$ to the multivariate setting.

Lastly, Theorem 6 in the appendix shows identification in the case where $Z$ is absolutely continuous under weaker assumptions on the supports $X_z$, but requiring $m$ to be a measure preserving $C^1$-diffeomorphism instead of simply being a measure preserving isomorphism, which is much stronger. Its statement is a straightforward generalization of the result in Torgovitsky (2015b).

To conclude this section, we want to stress again that Theorem 3 and its absolutely continuous counterpart from the appendix are not constructive identification results in the sense that they do not provide us with the function $m$. They just provide the identified set $I$ which we prove to contain a single element $m$. There are ways to estimate the function $m$ semi-parametrically like Komunjer & Santos (2010) or Torgovitsky (2016), but a fully nonparametric approach is still lacking. This is especially important to keep in mind in the following section where we show identification of the Hedonic model in multiple markets.

4 Applications: Engel curves, the BLP-, and Hedonic models

In order to showcase the applicability of Theorems 1, 2, and especially 3, we apply them in three different settings. First, we apply Theorems 1 and 2 in order to estimate quantile structural functions of Engel curves from the 1996 UK Family Expenditure survey analogous to Imbens & Newey (2009). Second, we apply Theorem 3 to the BLP (Berry, Levinsohn & Pakes 1995) model. Third, in our main application we apply Theorem 3 to Hedonic models with multivariate heterogeneity and endogenous characteristics, providing the first identification result in this setting and answering an open question posed in Chernozhukov, Galichon, Henry & Pass (2014) in the process. Let us start with the former.

4.1 Engel curves

Since Theorems 1 and 2 are direct generalizations of the control variable approach from Imbens & Newey (2009), we apply them to the same setting and estimate quantile structural functions of Engel curves from the 1995/1996 UK expenditure survey. Analogous to Blundell, Chen & Kristensen (2007) and Imbens & Newey (2009), $Y$ will be the share of expenditure on a commodity and $X$ will be the log of total expenditure. The instrument we use is gross earnings of the head
of the household, which was used in both Blundell, Chen & Kristensen (2007) and Imbens & Newey (2009). We estimate the quantile structural function when $Y$ is the share of expenditure on either food or leisure. Analogous to the argument of Imbens & Newey (2009), we can interpret the quantile structural function as giving quantiles of individual Engel curves across heterogeneous individuals as long as $\varepsilon$ exclusively represents heterogeneity and not other randomness like measurement error.

We use a subset of married and cohabiting couples where the head of the household is aged between 20 and 55, and couples with 3 or more children are excluded. We also exclude households where the head of the household is unemployed in order to have the instrument available. This leaves us with 1650 observations. Later, we will also include the number of children as a covariate. Since Theorem 2 requires absolutely continuous random variables, we cannot include the number of children directly; instead, we will restrict our data set further to families with at most one child and consider the child’s age as our second endogenous variable. For our illustrative purposes, we believe that this captures similar properties as the number of children. Restricting our sample to families with at most one child leaves us with 757 observations.

Before estimating the quantile structural functions for multiple endogenous variables, however, let us show that Theorem 1 is already very relevant in practice. Consider Figure 9 which depicts level sets of the conditional density $\hat{f}_{\hat{V}|X}$ for the case where the control variable $V$ is uniformly distributed on $[0,1]$, where $\hat{V}$ is estimated with bandwidths obtained via cross-validation, using the np package in R (Hayfield & Racine 2008). The large support assumption $V_x = V$ for all $x$ is badly violated, so that Proposition 2 cannot be applied directly. In their application Imbens & Newey (2009) assume that the large support assumption is satisfied.

![Figure 9: Level curves of $\hat{f}_{\hat{V},X}(v,x)$ (left) and $\hat{f}_{\hat{V}|X=x}(v)$ (right) for uniformly distributed $\hat{V}$](image)

However, Theorem 1 shows that one can construct $\hat{V}$ using any absolutely continuous dis-
tribution. We found that the large support assumption is close to being satisfied if \( \hat{V} \) follows a mixture of Normal distributions, in particular a mixture of 9 different Normal distributions, supported on \([-4, -3, -2, -1, 0, 1, 2, 3, 4]^T\) and with individual standard deviations of \([0.75, 1, 1, 1, 1, 1, 1, 0.75]^T\). Consider Figure 10. In this case \( \mathcal{V} \) is infinite in theory, but the practical construction of \( \hat{V} \) shows that the support is almost perfectly contained in \([-5, 5]\).

Figure 10: Level curves of \( \hat{f}_{\hat{V},X}(v, x) \) (left) and \( \hat{f}_{\hat{V}|X=x}(v) \) (right) for \( \hat{V} \) following a mixture of Normals

It turns out that the large support assumption is still violated in this case, but only very slightly and by far not as badly as in the case where \( V \) is uniformly distributed, so that the bounds we obtain are very tight. In fact, computing bounds analogously to Imbens & Newey (2009) via

\[
\hat{P}(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} [\hat{V}_i \neq \hat{V}_x],
\]

where

\[
\hat{V}_x := \{ V : \hat{f}_{V|X=x}(v) \geq \lambda_n, V \in \hat{V} \}
\]

is an estimator of the conditional support, \( \lambda_n > 0 \) is a (small) trimming parameter (in our case \( 10^{-4} \)), and \( \hat{V} \) is an estimator of the unconditional support, we obtain the results depicted in Figure 11. To interpret this figure, note that we obtain point identification at those \( x \in \mathcal{X} \) for which \( \hat{P}(x) = 0 \). As in the original application of Imbens & Newey (2009) the bounds get very large, especially at the boundaries of the support, when we choose \( V \) to follow a uniform distribution. In contrast, for our mixture distribution the bounds are actually very narrow, so that the estimators of the Engel curves we obtain are likely rather accurate.

Let us now estimate the quantile structural functions for this case to check if one obtains
different functional forms in those two cases. For this, recall that by Proposition 2 one can identify the quantile structural function by replacing \( \lambda(Y) = \mathbb{I}[Y \leq y] \) to obtain

\[
\int F_{Y|X,V}(y, x, v) F_V(dv) = E(E(\mathbb{I}[Y \leq y]|X = x, V))
\]

\[
\int \mathbb{I}[m(x, e) \leq y] F_\epsilon(de) = P(m(x, \epsilon) \leq y),
\]

so that the quantile structural function is

\[
q(\tau, x) = \left( \int F_{Y|X,V}(\tau, x, v) F_V(dv) \right)^{-1},
\]

completely analogous to Imbens & Newey (2009). Here \( \tau \in [0, 1] \) denotes the respective quantile. The estimator is simply a plug-in, so that

\[
\hat{q}(\tau, x) = \left( \frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y|X,V}(\tau, x, \hat{V}_i) \right)^{-1}.
\]

We estimate the conditional distribution with standard kernel estimates, unlike Imbens & Newey (2009) who use local linear regression. Local linear regression behaves better at boundaries of the support, but we show that with the tight bounds induced by a \( V \) which follows the above mixture of Normals, we still obtain very reasonable results, which are very similar to the results displayed in Imbens & Newey (2009). The estimated quantile structural functions for food expenditures we obtain using uniformly distributed \( V \) or \( V \) following the mixture of Normals.

![Figure 11: Bounds of the point-identification result for uniformly distributed \( V \) (thin line) and \( V \) following a mixture of Normals (thick line)](image)
described above are basically identical. Both do give the correct form of Engel curves, as they should be sloping downward, see Figure 12. The three different lines give quantile structural functions for the quartiles.

Figure 12: Estimations of the QSF of Engel curves for food expenditures, for uniformly distributed $V$ (left) and $V$ following the mixture of Normals (right) at quartiles

However, the estimates of Engel curves we obtain for leisure differ rather drastically, depending on which control variable we use, especially for higher expenditures, see Figure 13. In fact, for uniformly distributed $V$, the curves we estimate do not appear to be correct. Economic theory predicts an increasing functional form for Engel curves of leisure, which is the functional form we obtain when $V$ distributed as the mixture of Normals. In this case the results are very similar to the results obtained by Imbens & Newey (2009); moreover, since the bounds for identification are rather narrow, we believe that our point-identification result is more robust and credible than the original one. Theorem 1 is hence important to draw the correct conclusions from the data.

It is interesting to see that for the highest quartile the expenditures for leisure rise sharply for higher income echelons, a reasonable result which is still somewhat surprising in its magnitude. For the lowest quartile there is not a stark difference in terms of expenditures on leisure between the different levels of income, which is also to be expected.

The above results do not distinguish between households with children and households without children, however. The number of children is endogenous for demand if one does not follow the “purist” view that children should be treated like any other good, so that it should be included as an independent variable for demand estimation in this case (Browning & Lechene, 2003). Among the proposed instruments from the literature for this setting are age of the mother and educational attainment of both father and mother (Schultz, 1978), nonlinear functions of mother’s age and educational attainment (Mroz, 1987), and profession of the mother and her parents (Browning...
Figure 13: Estimations of the QSF of Engel curves for leisure expenditures, for uniformly distributed $V$ (left) and $V$ following the mixture of Normals (right) at quartiles

As our only instrument for the number of children we choose the age of the mother, which we consider to be continuous. The main reason is that we will have to compute an optimal transport map between $F_{X|Z=z}$ and $F_{V}$ for every realization $z \in Z$, and the more instruments we include for the number of children, the more realizations we have to loop over, as we would need to take all possible interactions between the two instruments.\(^{20}\)

As for the number of children, recall that Theorem 2 requires all endogenous variables to be continuous. We hence have two instruments, age of the mother as well as income of the head of the household, for the two endogenous variables log(total expenditure) and age of the child, as required by Theorem 2. As our control variable $V$ we use another mixture of Gaussians, this time a mixture of four bivariate Normal distributions:

$$V \sim \mathcal{N}
\begin{pmatrix}
0 \\
-2
\end{pmatrix},
\begin{pmatrix}
1.5 & 0 \\
0 & 0.45
\end{pmatrix}
+ \mathcal{N}
\begin{pmatrix}
1 \\
1
\end{pmatrix},
\begin{pmatrix}
1.1 & 0 \\
0 & 1.3
\end{pmatrix}
+ \mathcal{N}
\begin{pmatrix}
-0.5 \\
1
\end{pmatrix},
\begin{pmatrix}
0.9 & 0 \\
0 & 1.4
\end{pmatrix}
+ \mathcal{N}
\begin{pmatrix}
-1 \\
1
\end{pmatrix},
\begin{pmatrix}
0.75 & 0 \\
0 & 1.5
\end{pmatrix}$$

Figure 14 depicts a contour plot of this control variable.

In order to calculate the transport map $g^{-1}(x, z)$, we use and adapt the program developed in Gunsilius & Schennach (2017), which is rather intensive computationally as we have to obtain an optimal transport map between the random variables $X|Z = z$ and $V$ for every realization

\(^{20}\)Note that we do not take interactions between instruments for different endogenous variables, only for instruments of the same endogenous variable.
Figure 14: Contour plot of the density function used as a control variable in the multivariate case $z \in \mathcal{Z}$, which are 757 in our data set.\(^{21}\) Figures 15 and 16 depict the estimated quantile structural functions in this case.

Note that the estimates are somewhat crude. We have estimated the quantile structural function at only 16 values of log(total expenditure) as well as at age levels 0 to 15 in order to cut computational complexity.\(^{22}\) Also note that the functional forms of the quantile structural functions are as predicted by Economic theory. Interestingly, there is virtually no difference in the relative expenditure on food for the different age levels, or better: the differences in relative expenditure on food are so small that our relatively crude estimation does not pick them up. Leisure expenditures, on the other hand, are increasing with the age of the child, except for bumps from age 0 (i.e. married couples without children) to children of age 1, where the former seem to spend slightly more on leisure at all quartile levels.

Throughout, the levels of the quantile structural functions are not as high as in the univariate case, especially for leisure expenditures. The most likely explanation for this is that the large support assumption is not as close to being satisfied by the current control variable as in the univariate case. Moreover, the approach for obtaining the quantile structural functions involves many complex computations, and we have been using rather coarse grids whenever possible, which could be another reason for the discrepancy. Lastly, it is also debatable whether mother’s age is a good instrument for child’s age in this setting. The estimates in Figure 15 and 16 should hence be regarded as qualitative instead of quantitative. It would be interesting to see estimates for finer grids, different (and more) instruments, as well as different endogenous variables. We leave

\(^{21}\)For further information about the program we use to calculate the optimal transport map, we refer to Gunsilius & Schennach (2017).

\(^{22}\)In the UK FES adult age is defined to be 16.
Figure 15: Quantile structural functions for expenditure on food at the 25th (top left), 50th (top right), and 75th (bottom) percentile
Figure 16: Quantile structural functions for expenditure on leisure at the 25th (top left), 50th (top right), and 75th (bottom) percentile.
this for future research. Let us now to turn to applications of Theorem 3. We start with the BLP model.

4.2 The BLP model

The BLP model (Berry, Levinsohn & Pakes 1995) was introduced for identifying and estimating utility- and cost functions of participants in demand and supply systems of differentiated product markets when only aggregate market share data are available to the researcher. The only article providing results on identification of the BLP model to date is the seminal Berry & Haile (2014). In this article the authors need to make an artificial and rather strong index restriction, because their identification result relies on univariate identification results from the literature, in particular the identification result in Chernozhukov & Hansen (2005). Using Theorem 3, we can generalize their result directly to prove nonparametric identification without the need for the index restriction, complementing their result. Let us start with the demand side.

Demand side The demand side in this model is obtained by aggregating a continuum of individual discrete choice models in the following way, where we adapt the notation from Berry & Haile (2014). Each consumer $i$ in market $t$ chooses a good $j$ from a market $J_t := \{0, 1, \ldots, J_t\}$, which consists of a continuum of consumers with total measure $M_t$. A market is formally defined by $(J_t, \chi_t)$ with $\chi_t := (x_t, p_t, \xi_t)$. Here, $x_t = (x_{1t}, \ldots, x_{J_t})$ is a $K \times J_t$ matrix containing the observed and exogenous characteristics of the products in the market. $\xi_t := (\xi_{1t}, \ldots, \xi_{J_t})$ contains all of the unobservable characteristics at the product or market level and $p_t := (p_{1t}, \ldots, p_{J_t})$ contains observable endogenous characteristics, i.e. those characteristics which are correlated with $\xi_t$ like the price.

Consumer preferences in the BLP model are determined by indirect utilities in the sense that consumer $i$ in market $t$ has conditional indirect utilities $v_{i0t}, \ldots, v_{iJ_t}$. Following Berry & Haile (2014), we normalize the outside option $v_{i0t}$ to be zero, i.e. $v_{i0t} = 0$ for all $i$ and $t$ and assume that the utilities are independent and identically distributed across consumers and markets with joint distribution function $F_v(v_{i1t}, \ldots, v_{iJ_t}|\chi_t)$. Then the standard assumption is that $\arg\max_{j \in J_t} v_{ijt}$ is unique with probability 1, which leads to the following definition of the market shares $s_{jt}$ for each product $j$ in market $t$:

$$s_{jt} = \sigma_j(\chi_t) = P\left(\arg\max_{k \in J} v_{ikt} = j | \chi_t \right), \quad j = 0, \ldots, J,$$

under the normalization $s_{0t} = 1 - \sum_{k=1}^{J} s_{kt}$. Now here is where Berry & Haile (2014) are forced to introduce a strong assumption, the index restriction, which enables them to write the demand function element-wise for every $j$. In particular, they define a univariate index $\delta_{jt} = \delta_j(x_{jt}, \xi_{jt})$ for each product $j$, where $\delta_j$ is a function which is strictly increasing and continuous in the
The idea then is to write the demand function \( \sigma_j(\chi_t) \) for each \( j \) only in terms of \( x_{jt}, \xi_{jt}, \) and \( p_t, \) element-wise for every \( j, \) and then identify \( \xi_{jt} \) by inverting \( \sigma_j \) as well as the index \( \delta_j \) to get
\[
\xi_{jt} = \delta_j^{-1} \left( \sigma_j^{-1}(s_t, p_t), x_{jt} \right),
\]
requiring both \( \sigma_j \) and \( \delta_j \) to be strictly increasing and continuous functions in \( \xi_{jt}. \)

It is exactly here where the reader should make the connection to the introduction of this article. In fact, the functional form assumption of strict monotonicity and continuity on the index \( \delta_j \) and the demand function \( \sigma_j \) is the standard assumption from \cite{Matzkin2003}, and simply serves as a tool in order to work with a univariate unobservable for every distribution. Using Theorem 3, we are able to prove identification of the model without being forced to make any artificial assumptions index restrictions, therefore complementing the result in \cite{BerryHaile2014}. We do so as follows.

Firstly, we allow for a general demand function \( \sigma \) solving the demand problem (11) over all products \( j = 1, \ldots, J \) in the market simultaneously and for all individuals \( i \) in the continuum \( M_t, \) so that \( s_t = \sigma(\chi_t) \). This is the main difference to the index restriction: \cite{BerryHaile2014} allow for multiple products like we do, but they only do so element-wise, i.e. they treat every good separately. We on the other hand can allow for general interactions between the products. The outside option is still fixed as above. Analogous to \cite{BerryHaile2014}, we assume that this maximization problem has a unique solution. Note that this is the uniqueness assumption we need to make \( \sigma \) determinable. In addition we have to assume that \( \sigma \) is invertible between \( \xi_t \) and \( s_t, \) a condition which might be hard to satisfy in practice, but has been the standard assumption in this literature (see \cite{Matzkin2007a} for further discussion of this point). \cite{BerryHaile2014} require monotonicity and continuity in every element, which is a sufficient condition for invertibility of \( \sigma \) and might be even harder to satisfy in practice.

Measure preservation is a natural assumption for BLP models. Recall that \( s_t \) is the vector of market shares of every product, which possesses a certain (conditional) probability distribution \( P_{s_t|x_t,p_t}. \) This distribution is just the distribution of choices of individuals \( i \) in the market \( t, \) so that one can view each point in the support of \( s_t \) conditional on \( x_t \) and \( p_t, \) denoted by \( S_{x_t,p_t}, \) as the purchase plan of an individual \( i, \) determining the probability with which this individual is to buy which product \( j \) in the market. Then this individual \( i \) needs to have a certain evaluation of the products \( j = 1, \ldots, J, \) which is unobservable to the Econometrician, i.e. a distribution over the unobservables \( x_{jt} \) analogous to the purchase plan of the individual; this can be thought of as giving for each product \( j \) a probability of how “important” the respective unobservable \( x_{jt} \) of the product is for the individual’s choice.

Since all individuals lie on a continuum, it makes more sense to talk about sets of individuals instead of unique individuals. Therefore, every (Borel-) set \( E \in S_{x_t,p_t} \) of individuals with purchase

---

\footnote{In the main text, they even assume that \( \delta_j \) is linear, a much stronger assumption, but they relax this assumption to allow for strictly increasing and continuous \( \delta_j \) in the appendix, and this is the result we focus on.}
plans $P_{s_t|x_t,p_t}(E)$ must have a corresponding set in $P_{t|x}$ which is of the same size, because all evaluations and purchase plans are based on the same set of individuals. But this is exactly the definition of a measure preserving demand function $\sigma$, i.e. we require

$$P_{s_t|x_t,p_t}(E) = P_{t|x} \left( \sigma^{-1}(x_t,p_t,E) \right) \quad \text{for all} \quad E \in \mathcal{B}_{s_t|x_t,p_t}. \quad (12)$$

Now, again, as in the example of Hedonic models, the need for Theorem 3 arises from the fact that the $p_t$ are endogenous. To the best of our knowledge, this provides the first instance where nonseparable triangular models can be used for identification of the BLP model. Those results were not possible previously, because they required that the second stage in those models be univariate for point-identification, as argued in Berry & Haile (2014, p. 1754). Providing complete identification of the BLP model therefore provides another instance proving how important a multivariate generalization of these identification results really is. In order to apply Theorem 3 we need to model the first stage relationship

$$p_t = h(z_t,u_t), \quad (13)$$

where $z_t$ is a set of instruments which are excluded from the demand model, $u_t$ is a vector of unobservables of the same dimension as $p_t$ and $h$ is a measure preserving isomorphism which we do not specify more closely. As in the Hedonic model example, we apply the multivariate control variable approach, so that we cannot assume more on $h$ than that it is some measure preserving isomorphism. Again, an alternative would be to assume a distribution for $u_t$ and require $h$ to be the Brenier map between $u_t$ and $p_t$.

In our setting it is very natural to let $p_{jt}$ be multivariate for every $j$, hence letting $p_t$ be a matrix. All we have to do to make this work is to vectorize the matrix $p_t$ by stacking each column onto one another, i.e. identifying the matrix space $\mathbb{R}^{J_t \times K}$ with $\mathbb{R}^{J_t \cdot K}$, where $K$ is the number of columns for every $p_{jt}$. Note again, that we can allow for instruments $z_t$ which are discrete and lower dimensional than the endogenous variables $p_t$, allowing for binary policy changes. All the instruments need to satisfy is $z_t \perp (u_t, \xi_t)$ and that they have an influence on each element of $p_t$. Let us now state the identification result for the demand side. The observables of the market are $(M_t,x_t,p_t,s_t,z_t)$. The demand side is modeled through (12) and (13). Then the following holds.

**Proposition 3** (General identification of the demand side in the BLP model). *In the case where the instruments $z_t$ are absolutely continuous, let the regularity assumptions hold as stated in Theorem 6. In the case where the instruments $z_t$ are discrete, let the first stage (13) only be specified to be a measure preserving isomorphism, let the demand function $\sigma$ satisfy Assumption 7 and let Assumptions 5 – 9 hold for $p_t$, $z_t$, $u_t$, and $\xi_t$. Then the model (12) and (13) is identified in the sense that the identified set

$$I := \{ \sigma \in \mathcal{H}(S_{x_t,p_t}, \xi_t) : (\sigma^{-1}(x_t,p_t,s_t),V) \perp z_t \}$$

37
contains an almost everywhere unique element $\sigma$. Here, $V$ is a control variable of the same dimension as $p_t$ as in Theorem 2 and $\mathcal{H}(S_{x_t,p_t}, \xi_t)$ is the set of all isomorphisms between $\xi_t$ and $s_t$ for exogenous $x_t$ and $p_t$.

The proof of this proposition follows immediately from Theorem 3 or Theorem 6. Proposition 3 therefore provides nonparametric identification for the demand side of the BLP model in the most general case, only requiring $\sigma$ to be a measure preserving isomorphism. Note that Assumption 7 requires a normalization of the demand function. This can be done by assuming a multivariate uniform distribution for $\xi_t$, in the sense that all $\xi_{jt}$ are uniformly distributed, which is the analogue to the normalization in Berry & Haile (2014) who assume a univariate uniform distribution for every $\xi_{jt}$. In our case, one is actually free to model the dependency structure between the $\xi_{jt}$, however, i.e. one is not required to assume that they are all independently distributed. As for $\sigma$ being a measure preserving isomorphism, this is satisfied as soon as the utility maximization problem has a unique and invertible solution.

The nice thing about Proposition 3 is that the assumptions of a unique and measure preserving demand function $\sigma$ are natural and can be implied by the set-up of the model. Also notice how our approach allows for multivariate $p_t$ even from the set-up. The last interesting and also important thing to recall is that we can allow for instruments to be of lower dimension than the endogenous variables $p_t$. This is especially important in practice. In fact, it might often be the case that there is a dichotomous shock introduced into the model, possibly through a policy change, which can serve as an instrument. If this policy change is truly independent and is such that it influences all $p_t$, then it alone can serve as a single instrument to identify the whole demand side of the model, under the restriction that Assumption 9 on the supports of $P_{p_t|z_t=z}$ and $P_{p_t|z_t=z'}$ is satisfied; but this assumption can be checked in higher dimensions and simply be eyeballed in the case where $p_t$ is two-dimensional, an important special case.

Supply side  Having identified the demand side, one can model the supply side in basically two ways. First, one can simply assume that one knows the oligopolistic structure of the supply side in which case one can immediately deduce the vector of marginal costs $mc_t := (mc_{1t}, \ldots, mc_{Jt})$ by

$$mc_t = \psi(s_t, M_t, \sigma, p_t),$$  \hspace{1cm} (14)

since all quantities on the right hand side are observed ($s_t, M_t, p_t$) or identified ($\sigma$). Note that (14) is more general than the function proposed in Berry & Haile (2014), which is, again, only defined element-wise, i.e. one $\psi_j$ for every product $j$, analogous to their element-wise definition of the demand function $\sigma_j$. In the case for known $\psi$, there is nothing to do from an Econometric perspective, as one simply assumes away the problem of identifying the respective monopoly structure, i.e. the function $\psi$. This can be warranted in some cases, where one has additional
knowledge on the oligopoly structure. Based on this, one can identify the cost functions

\[ mc_t = c(Q_t, w_t, \omega_t) \]  

with some instrument (i.e. supply shifter) and Theorem 3 or Theorem 6. Here, \( w_t := (w_{1t}, \ldots, w_{Jt}) \) are observable and exogenous cost shifters, and \( \omega_t := (\omega_{1t}, \ldots, \omega_{Jt}) \) are unobservable cost shifters. Let \( J_J \) denote the set of products produced by the firm producing product \( j \). Let \( q_{jt} = M_t s_{jt} \) be the quantity produced of good \( j \) in equilibrium and let \( Q_{jt} \) be the vector of quantities of all goods \( k \in J_J \). This setting is completely analogous to the demand side if we replace \( w_t \equiv x_t, \omega_t \equiv \xi_t, \) and \( Q_t = p_t. \)

The important and more realistic way to model the supply side, however, is to allow for an unknown function \( \psi \). Note that in this case, there are several approaches towards identification. In principle, there are four things to identify in the model: \( mc_t, \psi, c, \) and the unobservable shocks \( \omega_t. \) The approach in Berry & Haile (2014) is to combine (14) and (15) into one equation

\[ \omega_t = c^{-1}(Q_t, w_t, \psi(s_t, M_t, \sigma, p_t)) := \pi^{-1}(p_t, M_t, s_t, w_t), \]  

(16)

eliminating \( mc_t \) in the process. This approach enables them to identify the unobservable shocks \( \omega_t \) as well as the function \( \pi \) which incorporates both \( c \) and \( \psi. \) Their approach consists of assuming that both \( c \) and \( \psi \) can be written element-wise, with each \( c_j \) being linear in \( w_{jt} \) and \( \omega_{jt}. \) We can do the same but much more generally with our approach, simply requiring \( \psi \) and \( c \) to be measure preserving isomorphisms. Then \( \pi^{-1} \) must be a measure preserving isomorphism, too. Now realize that (16) is perfectly analogous to (12). Therefore, we can apply Theorem 3 again in order to establish identification. We only need some instrument \( z_t \) for the supply side. Then a first stage for the endogenous \( p_t \) is

\[ p_t = h(z_t, u_t), \]  

(17)

where, again, \( u_t \) is unobservable and of the same dimension as \( p_t \) and \( h \) is a measure preserving isomorphism, exactly as in (13). Then we can state the following identification result for the supply side, which we only state in terms of Theorem 3—the statement for Theorem 6 is of course analogous.

**Proposition 4.** In the case where the instruments \( z_t \) are discrete, let the first stage (13) be known to be a measure preserving isomorphism, let \( \pi \) satisfy Assumption 7 and let Assumptions 5 – 9 hold for \( p_t, z_t, u_t, \) and \( \omega_t. \) Then the model consisting of (16) and (17) is identified in the sense that the identified set

\[ \mathbb{I} := \{ \pi \in \mathcal{H}(S_{w_t,M_t,p_t}, \omega_t) : (\pi^{-1}(w_t, M_t, p_t, s_t), V) \perp z_t \} \]

contains an almost everywhere unique element \( \pi. \) Here, \( V \) is a control variable of the same dimension as \( p_t \) as in Theorem 3 and \( \mathcal{H}(S_{w_t,M_t,p_t}, \omega_t) \) is the set of all isomorphisms between \( \omega_t \)
and \( s_t \) for exogenous \( w_t \) and \( p_t \) and fixed \( M_t \).

The proof again follows immediately from Theorem 3. Again, one has to make a normalization assumption, but assuming that \( mc_t \) is multivariate uniform is not a strong restriction.

Now, in many cases one is actually interested in identifying both functions \( c \) and \( \psi \) separately, because \( \psi \) gives information about the oligopoly structure of the market. This can again be done in our setting if one has proper instruments for both equations, which in general is not a strong requirement, because one can use standard exogenous shifters as argued in Berry & Haile (2014).

The idea is to use our identification approach first on (14) and then on (15) once we have identified \( mc_t \). Let us assume that there are appropriate sets of shifters (i.e. instruments) \( z^1_t \) and \( z^2_t \) for (14) and (15), respectively. Note that we do admit the likely case \( z^1_t = z^2_t \). We also need two first stage equations:

\[
p_t = h_1(z^1_t, u^1_t),
\]
(18)

and

\[
Q_t = h_2(z^2_t, u^2_t),
\]
(19)

for unobservables \( u^1_t \) and \( u^2_t \) with the same dimension as \( p_t \) and \( Q_t \), respectively, and \( h_1 \) and \( h_2 \) measure preserving isomorphisms. Since \( Q_t \) is a matrix, we rely on the simple trick of writing it as a vector, stacking the columns upon one another, as mentioned. The main identification result for the supply side is then as follows. Again, we only state the result for Theorem 3, but the result for Theorem 6 is perfectly analogous.

**Proposition 5** (General identification of the supply side of the BLP model). In the case where the instruments \( z^1_t \) and \( z^2_t \) are discrete, let the first stages (18) and (19) only be known to be measure preserving isomorphisms, let \( c \) and \( \psi \) satisfy Assumption 7 and let Assumptions 5 – 9 hold for \( p_t, Q_t, z^1_t, z^2_t, u^1_t, u^2_t, mc_t \), and \( \omega_t \). Then the two models consisting of (14) and (18) as well as (15) and (19) are identified in the sense that the identified sets

\[
\mathbb{I}_1 := \{ \psi^{-1} \in \mathcal{H}(S_{w_t,M_t,p_t}, mc_t) : (\psi(\sigma, M_t, p_t, s_t), V_1) \perp z^1_t \}
\]

and

\[
\mathbb{I}_2 := \{ c \in \mathcal{H}(S_{w_t,M_t,Q_t}, \omega_t) : (c^{-1}(w_t, M_t, Q_t, s_t), V_2) \perp z^2_t \}
\]

contain almost everywhere unique elements \( \psi \) and \( c \), respectively. Again, \( V_1 \) and \( V_2 \) are control variables of the same dimension as \( p_t \) and \( Q_t \), respectively, and \( \mathcal{H}(S_{w_t,M_t,p_t}, mc_t) \) and \( \mathcal{H}(S_{w_t,M_t,Q_t}, \omega_t) \) are the respective sets of homeomorphisms.

**Proof.** First, one needs to identify \( mc_t \) and \( \psi \) in (14) and (18). This follows immediately from Theorem 3 and the fact that \( \sigma \) has already been established to be identified on the demand side by Proposition 3. Then once \( mc_t \) is identified, one can turn to the identification of \( c \) and \( \omega_t \) in (15) and (19), which also follows immediately from Theorem 3 and the fact that \( mc_t \) has already been identified.
Remarks and outlook Propositions \(\text{\textsuperscript{3}}\) and \(\text{\textsuperscript{5}}\) together establish general identification of the BLP model, without the need to make artificial index restriction assumptions and hence allowing for the most general set-up. The main assumptions, in addition to regularity assumptions like continuity and invertibility, guaranteeing this result are measure preservation and uniqueness of the demand function \(\sigma\) as well as the cost functions \(\psi\) and \(c\). Both are very natural and fundamental, being a direct result of the model set-up. As a result of the endogeneity of quantity and price, we rewrote both sides of the model as a nonseparable triangular system and applied our two results from the previous section to guarantee point identification of the respective functions. There are certainly other ways for identification of this model than using Theorem \(\text{\textsuperscript{3}}\) or Theorem \(\text{\textsuperscript{6}}\), but both are very general and it is rather unlikely that even more general Theorems will hold for this setting. Note that this application was just an outline that a different identification result to Berry & Haile (2014) holds, where one does not have to make index restriction assumptions or assume that the demand function is element-wise strictly increasing and continuous, but can allow for more general results. The discussion in this section was very high level. A next step from here is to find appropriate low level assumptions on the demand function (other than monotonicity) which implies that it is measure-preserving and invertible. Those should come from Economic Theory. Let us now turn to the main application of our main result where we prove nonparametric identification of Hedonic models with multivariate unobservable heterogeneity and endogenous characteristics.

4.3 Hedonic models with endogenous characteristics

Theorem \(\text{\textsuperscript{3}}\) enables us to prove nonparametric identification of general Hedonic models with multivariate unobservable heterogeneity and endogenous observable characteristics in a multi-market setting, the main result of this whole section.

Seminal identification results in Ekeland, Heckman & Nesheim (2004) and Heckman, Matzkin & Nesheim (2010) have focused on single-market Hedonic models with univariate unobservable heterogeneity and stronger functional form assumptions on the utility function. The recent working paper Chernozhukov, Galichon, Henry & Pass (2014) extended these results to single-market Hedonic models with multivariate unobservable heterogeneity using optimal transport theory. We complement these results in this subsection by for the first time allowing for endogenous observable characteristics in a multi-market setting under multivariate unobservables.

We adapt the notation in Chernozhukov, Galichon, Henry & Pass (2014) slightly to make it compatible with the notation from the previous section. We consider an environment where consumers and producers trade a good or contract which is fully characterized by its type or quality \(y \in \mathcal{Y} \subseteq \mathbb{R}^d\). Its price, \(p(y)\) is determined endogenously in equilibrium. Producers \(\tilde{W}\) and consumers \(\tilde{X}\) are characterized by their respective types, \(\tilde{w} \in \tilde{W} \subseteq \mathbb{R}^{m+d}\) and \(\tilde{x} \in \tilde{X} \subseteq \mathbb{R}^{k+d}\). They are price takers and maximize quasi-linear utility \(U(\tilde{x}, y) - p(y)\) and profit \(p(y) - C(\tilde{w}, y)\), respectively, where \(U\) is upper semicontinuous and bounded and \(C\) is lower semicontinuous and
bounded. Both are normalized to zero in case of nonparticipation. We use the following equilibrium concept.

**Assumption 10** (Equilibrium concept). The pair \((\gamma, p)\), for \(\gamma\) a probability measure on \(\bar{X} \times \mathcal{Y} \times \hat{W}\) and \(p\) a function on \(\mathcal{Y}\), is a hedonic equilibrium in the sense that \(\gamma\) has marginals \(P_{\bar{X}}\) and \(P_{\hat{W}}\) and for \(\gamma\)-almost all \((\bar{x}, y, \hat{w})\)

\[
U(\bar{x}, y) - p(y) = \max_{y' \in \mathcal{Y}} (U(\bar{x}, y') - p(y)),
\]

\[
p(y) - C(\hat{w}, y) = \max_{y' \in \mathcal{Y}} (p(y) - C(\hat{w}, y')).
\]

In addition, observed qualities \(y = y(\bar{x}, \hat{w})\), which maximize the joint surplus \(U(\bar{x}, y) - C(\hat{w}, y)\) for all \((\bar{x}, \hat{w}) \in \bar{X} \times \hat{W}\), lie in the interior of \(\mathcal{Y}\).

For absolutely continuous distributions \(P_{\bar{X}}\) and \(P_{\hat{W}}\) [Ekeland (2010) and Chiappori, McCann & Nesheim (2010)] show that a pure equilibrium exists and is unique under the twist condition from optimal transport (Villani 2008, p. 216) which requires that the gradients \(\nabla_\bar{x} U(\bar{x}, y) := \frac{\partial}{\partial \bar{x}} U(\bar{x}, y)\) and \(\nabla_\hat{w} C(\hat{w}, y) := \frac{\partial}{\partial \hat{w}} C(\hat{w}, y)\) exist and are injective as functions of quality \(y\). To obtain an estimable model, unobservable heterogeneity is introduced in the following way\(^{24}\).

**Assumption 11** (Unobservable heterogeneity). The observable type \(\bar{x}\) consists of an observable portion \(x \in \mathbb{R}^k\) as well as a random unobservable portion \(\varepsilon \in \mathbb{R}^d\), i.e. \(\bar{x} = (x, \varepsilon)\).\(^{24}\) Furthermore, utility \(U(\bar{x}, y)\) can be decomposed as \(U(\bar{x}, y) = \bar{U}(x, \varepsilon) + \xi(x, \varepsilon, y)\).

In this setting, the object of interest for identification is the deterministic component of the utility function, \(\bar{U}(x, y)\). Denote \(V(x, y) = p(y) - \bar{U}(x, y)\). Here is where [Chernozhukov, Galichon, Henry & Pass (2014)] need to make the rather strong assumption that the observable characteristics \(X\) are exogenous, i.e. \(X \perp \varepsilon\). The idea for identification then proceeds roughly as follows. The Monge-Kantorovitch problem leads to natural functional form restrictions in this setting as the consumer’s optimization program is to choose \(y\) such that

\[
V^\gamma = \sup_y \{\xi(x, \varepsilon, y) - V(x, y)\}\] \(^{20}\) (20)

Identification of \(\bar{U}(x, y)\) requires identification of \(V(x, y)\); to identify the latter one can use two approaches. One way is to show that the pair \((V(x, y), V^\gamma(x, y))\) uniquely solves the dual problem of the Monge-Kantorovitch problem under the cost function \(\xi(x, \varepsilon, y)\) in the general case under some regularity assumptions. The second route to identify \(V(x, y)\) is via the optimal planner’s problem, which takes the form of the Monge-Kantorovitch problem with general cost function

\(\xi(x, \varepsilon, y) = y'\varepsilon,\) as in this case \(V^*(x, y) = \sup_{y' \in \mathcal{Y}} \{y' \varepsilon - V(x, y)\}\), where \(V^*(x, y)\) is the convex conjugate of \(V(x, y)\).

\(\text{\(^{24}\)We only focus on the consumer’s side throughout, because identification of the supplier’s side is analogous.}\)

\(\text{\(^{25}\)Note that \(\varepsilon\) is assumed to be of the same dimension as \(Y\).}\)

\(\text{\(^{26}\)Note that this optimization problem is the definition of the convex conjugate (Rockafellar 1997, §12) if} \quad \xi(x, \varepsilon, y) = y' \varepsilon, \text{ as in this case } V^*(x, y) = \sup_{y' \in \mathcal{Y}} \{y' \varepsilon - V(x, y)\}, \text{ where } V^*(x, y) \text{ is the convex conjugate of } V(x, y).\)
\(\xi(x, \varepsilon, y)\). In the special case \(\xi(x, \varepsilon, y) = y'\varepsilon\) considered in Ekeland, Heckman & Nesheim (2004) for example, the cost function for the Monge-Kantorovich problem is the squared Euclidean norm, in which case the inverse demand function \(m^{-1}(x, y)\) to take the form of the Brenier map as the determinable measure preserving map between \(\mathcal{P}_Y|X\) and \(\varepsilon\) under regularity assumptions. To root this set-up in our notation from above, note that \(m^{-1}(x, y)\) is the inverse of the measure preserving map \(m : \mathcal{E} \rightarrow \mathcal{Y}_x\) from the setting in section 3.

Allowing for general \(\xi(x, \varepsilon, y)\) is a generalization of the concept of convex duality, leading to (20), see Villani (2003, Chapter 2.4). (20) possesses a dual problem of the form

\[
(V^{\gamma})^{\gamma} = \sup_{\varepsilon} \{\xi(x, \varepsilon, y) - V^{\gamma}(x, y)\},
\]

analogous to convex duality. For standard convex duality, if \(V\) is convex and proper, it holds that \((V^*)^* = V\) (Rockafellar 1997, §12). Therefore, if the solution \((V^{\gamma})^{\gamma}\) to the dual problem (21) coincides with \(V\), we say that \(V\) is \(\gamma\)-convex. This idea is also used in Chernozhukov, Galichon, Henry & Pass (2014) in the following assumption, which we also require.

**Assumption 12.** \(V(x, y)\) is \(\gamma\)-convex.

Under some further regularity assumptions Chernozhukov, Galichon, Henry & Pass (2014) then use the fact that the Monge-Kantorovich problem has a unique measure preserving solution, so that \(m^{-1}(x, y)\) is determinable in the sense of Definition 3. Note that it is the assumption of exogenous \(X\) which enables them to identify \(m^{-1}(x, y)\), because they use the functional form restrictions imposed by the Monge-Kantorovich on \(m^{-1}(x, y)\) to identify it, generalizing the identification result from Matzkin (2003) in the process. They cannot allow for endogenous \(X\), however, and explicitly leave open the problem of identifying it. This is where we can apply Theorem 3.

For the endogenous case instruments in the form of exogenous shifters are required. One very convenient setting exists if the researcher has access to data in several, i.e. at least two, disjoint markets \(z\) and \(z'\). This is where Theorem 3 shines, because one can consider the two markets as realizations of a binary instrument \(Z\) under the assumption that \(\varepsilon\) does not change between markets, i.e. \(Z \perp \varepsilon\). Then if the distribution of \(X\) changes between the two markets, \(Z\) is an instrument for \(X\), as needed for Theorem 3. Formally, we have to include a first stage \(X = h(Z, \eta)\), where \(h\) satisfies Assumptions 3 and 4 for every market \(z_i\) and the unobservable \(\eta\) is of the same dimension as \(X\).

**Assumption 13 (Multiple markets as a discrete instrument).** The distribution of the unobservable \(\varepsilon\) does not change between different markets \(z, z'\), but the distribution of \(X\) does.

As stated above, there are two ways to work with and prove Theorem 3. One either assumes \(F_\eta\) is known and \(h\) is the Brenier map directly, as done in the proof of Theorem 3, or one uses the multivariate control variable approach and only assumes that \(h(z, \cdot)\) is a measure preserving isomorphism. For Hedonic models the control variable approach is more natural as it allows for
much more general functional form assumptions on $h$. We will go the latter path this time. Now in order to apply Theorem 3, we need to not only guarantee that $m^{-1}(x,y)$ is the determinable measure preserving map between $P_{Y|X=x,Z=z_{i}}$ and $P_{\varepsilon|X=x,Z=z_{i}}$, but also that it is continuous in all $x$. It turns out that under reasonably weak regularity assumptions we cannot guarantee this strong form of continuity; we can guarantee continuity in measure, however, so that we have to use Assumption 5'.

**Assumption 14** (Regularity assumptions). The following hold:

(i) $U(\tilde{x}, y)$ and $C(\tilde{y}, y)$ are differentiable with respect to $x$ and $w$, respectively, for all $\tilde{x}$ and $\tilde{w}$.

(ii) $\xi(x, \varepsilon, y)$ is continuously differentiable with respect to $y$ for all $x$, $\varepsilon$, and $y$.

(iii) $\det(\nabla_{\varepsilon}^{2}\xi(x, \varepsilon, y)) \neq 0$ for all $x, \varepsilon, y$, where $\nabla_{\varepsilon}^{2}\xi \equiv (\partial^{2}\xi/\partial \varepsilon_{i} \partial y_{j})_{ij}$ denotes the Hessian.

(iv) Twist condition: For all $x$ and $y$, the gradient $\nabla_{y}\xi(x, \varepsilon, y)$ of $\xi(x, \varepsilon, y)$ in $y$ is injective as a function of $\varepsilon$.

(v) $h$ is only known to be a measure preserving isomorphism between $\eta$ and $X$.

(vi) $P_{\varepsilon}$ is known.

(vii) $F_{X|Z=z_{i}}$ is absolutely continuous for all markets $z_{i}$ and satisfies Assumptions 5 and 9.

(viii) $P_{\varepsilon|X,Z}$ satisfies Assumption 8.

Parts (i) – (iv) of Assumption 14 guarantee existence and uniqueness of an optimal transport map between $P_{Y|X=x,Z=z_{i}}$ and $P_{\varepsilon|X=x,Z=z_{i}}$ for all $x$ and $z_{i}$ (see e.g. Chapter 10 in [Villani 2008]) and are the same as in [Chernozhukov, Galichon, Henry & Pass (2014)]. Parts (v) – (vii) are the assumptions we need to require for Theorem 3. Overall, our regularity assumptions are only slightly more demanding than the ones of [Chernozhukov, Galichon, Henry & Pass (2014)] in the exogenous case, but we can allow for endogenous characteristics. Of course, we need to require that $F_{X|Z=z_{i}}$ satisfy Assumption 9 which requires that their supports are convex, do not change between markets, and admit an appropriate manifold $M(z_{i}, z_{j})$. This leads to the following result.

**Theorem 4** (Identification of utility functions). In the Hedonic model defined in Assumptions 10 and 11 assume $\varepsilon \perp \perp X$. Let the researcher have access to data in at least two disjoint markets $z$ and $z'$ and assume a first stage of the form $X = h(Z, \eta)$ for $\eta, X \in \mathbb{R}^{k}$. Furthermore, let Assumptions 12, 13, and 14 hold. Then the utility function $\bar{U}(x, y)$ is identified for almost every $(x, y) \in X \times Y$ up to an additive constant.

**Intuition** The idea for the proof of Theorem 4 is to apply Theorem 3. As stated, the first four parts of Assumption 14 guarantee existence and uniqueness of the optimal transport map $m^{-1}(x,y)$ as proved in [Chernozhukov, Galichon, Henry & Pass (2014)], which is the determinable
measure preserving map between $P_Y|_{X=x, Z=z_i}$ and $P_Z|_{X=x, Z=z_i}$ required by Theorem 3. The additional requirement that this measure preserving map be continuous in measure for every $x$ is guaranteed by the existence of a continuous disintegration and a slight generalization of stability results for the Monge-Kantorovich problem (Villani 2008, Chapter 5). The other regularity assumptions then guarantee that $m^{-1}(x, y)$ can be identified for almost every $y$ and $x$ through Theorem 3. Assumption 12 is needed to guarantee differentiability of $V(x, y)$, which in turn leads to the identification of $U(x, y)$ for almost every $x$ and $y$ through a first order condition since $m^{-1}(x, y)$ is identified for almost every $x$ and $y$. The detailed proof is in the appendix.

Theorem 4 is the first result in the literature to use data from multiple markets to identify Hedonic models with multivariate unobservable and endogenous characteristics; it therefore answers the open question stated in Chernozhukov, Galichon, Henry & Pass (2014) who ask under which conditions one can identify Hedonic models with multivariate unobservables that are not independent of the observables.

5 Conclusion

In this article we have proposed the first nonparametric identification results for nonseparable triangular models with a multivariate first- and second stage. We have derived three main theoretical results.

The main result is a direct generalization of the seminal results from Torgovitsky (2015) and d’Haultfoeille & Février (2015) for point-identification of nonseparable triangular models with discrete instruments. In particular, we are the first to derive primitive conditions on the data generating process under which a nonseparable triangular model with a multivariate first and second stage is identified. This result is widely applicable. In fact, we use it to derive assumptions under which both the supply and the demand side of the BLP model are nonparametrically identified, even under general heterogeneity. Previously, one had to assume that all heterogeneity can be summarized in an index (Berry & Haile 2014), an artificial and strong restriction. As a second application, we prove the first nonparametric identification result for Hedonic models with endogenous characteristics and multivariate heterogeneity, treating different markets as realizations of a discrete instruments. In particular, this answers an open question in Chernozhukov, Galichon, Henry & Pass (2014) showing when Hedonic models with general heterogeneity are identified under endogeneity.

In addition we have provided a generalization of the control variable approach (Blundell & Powell 2003, Imbens & Newey 2009). In particular, we have shown that in nonseparable models with a univariate first stage, i.e. only allowing for one endogenous variable in the model, the control variable approach from Imbens & Newey (2009) can be generalized to allow for a wide variety of absolutely continuous distributions for the control variable, not just the uniform distribution. Interestingly, one does not need to make any stronger assumptions than those needed for a uniform control variable. This generalization is important in practice, because the large
support assumption needed for identification of average effects is often badly violated for uniformly distributed control variables, as in the application of estimating Engel curves considered in this article. We show that for a control variable whose distribution is a mixture of Normal distributions, the large support assumption is very close to being satisfied.

Furthermore, we generalize the control variable approach to allow for more than one endogenous variable in the model. This generalization is not straightforward since in the univariate case the result relies on the complete ordering of the real numbers. Therefore, we show that in order to allow for several endogenous variables in a truly nonseparable triangular model, one cannot make any functional form assumptions on the production function of the first stage besides the fact that it preserves measure and is invertible. In that case a multivariate generalization of the control variable approach to higher dimensions is straightforward and we illustrate it by estimating quantile structural functions of Engel curves for families with at most one child, where we consider total expenditure as well as the age of the child as endogenous variables.

We were able to obtain these results, because we use and derive some new results from the theory of optimal transport, in particular, we derive some apparently new results about properties of the Brenier map, the arguably most natural generalization of a strictly increasing and continuous function to the multivariate setting. In particular, we prove a new result on the dynamics of Brenier maps between two absolutely continuous measures whose supports coincide, and defining a criterion for a fixed set. This result builds the heart of our third main theoretical result, but is also of interest in itself as it provides a new explanation for how equilibria are formed in General Equilibrium theory.

Finally, our main theoretical identification result is non-constructive as it is a direct generalization of the seminal result of Torgovitsky (2015). There are ways to use this identification result for semi-parametric identification, but a fully nonparametric approach is still lacking. The next important step will hence be to find slightly stronger assumptions than the current ones which would enable nonparametric estimation and inference in these models. Moreover, since the identification result rely upon optimal transport theory, it is imperative to derive further statistical properties of these maps, in particular their large sample distributions.

References


Matzkin, R. (2007a), Heterogeneous choice. advances in economics and econometrics, theory and applications, in ‘Ninth World Congress of the Econometric Society’.


A The Monge-Kantorovich problem

A way to model determinable measure preserving maps is via the theory of Optimal Transport, a very active research area in (Applied) Mathematics with important recent applications in Economics and Econometrics. For references on this vast subject we refer to Gangbo & McCann (1996), Rachev & Rüschendorf (1998), Villani (2003), Villani (2008), and Santambrogio (2015); the recent addition to this literature, Galichon (2016), approaches the subject through the lens of Economics.

The original goal of this area of research is to find a measure preserving map \( T \) which transports one probability measure \( P_\varepsilon \) onto another probability measure \( P_Y \) in a “cost-efficient way”. The set up for this is the Monge-Kantorovich problem. To be precise, the Monge- and the Kantorovich problem are actually two different problems, the latter being the convex relaxation of the former. Monge’s Problem asks for an optimal transport map between two (probability) measures, \( P_\varepsilon \) and \( P_Y \), where optimality is measured with respect to some cost function \( c : \mathcal{E} \times \mathcal{Y} \rightarrow \mathbb{R} \).

\[
\min_{T : \mathcal{E} \rightarrow \mathcal{Y} \text{ measurable}} \int c(e,T(e)) \, dP_\varepsilon(e)
\]

\[\text{(22)}\]

Here \( y \in \mathcal{Y} \), \( e \in \mathcal{E} \), and \( \Pi(P_Y, P_\varepsilon) \) is the set of all probability measures on \( \mathcal{Y} \times \mathcal{E} \) such that the marginal distributions of some \( \pi \in \Pi(P_Y, P_\varepsilon) \) are precisely \( P_Y \) and \( P_\varepsilon \). The Kantorovich problem between (probability) measures \( P_Y \) and \( P_\varepsilon \) under some cost function \( c : \mathcal{Y} \times \mathcal{E} \rightarrow \mathbb{R} \) asks for an optimal transport plan in the sense that the transport does not have to be accomplished through a function as in the Monge problem, but is concentrated on the support \( \Gamma \) of a joint probability distribution \( \gamma \) which has \( P_\varepsilon \) and \( P_Y \) as marginals.

\[
\min_{\pi \in \Pi(P_Y, P_\varepsilon)} \int_{\mathcal{Y} \times \mathcal{E}} c(y,e) \, d\pi(y,e).
\]

\[\text{(23)}\]

To make the two problems more tangible, one can picture a pile of sand and some hole of the same volume. Then the Kantorovich problem asks for the most cost effective way to put the sand into the hole, allowing each grain of sand to be split up further. The Monge problem requires the optimal transportation plan mapping \( P_\varepsilon \) to \( P_Y \) to be such that no grain of sand be split, i.e. that the transport can actually be accomplished through some function \( T \). For many cost functions \( c \), the solution to the Monge and the Kantorovich problem actually coincide under the assumption that \( P_\varepsilon \) is absolutely continuous, so that it is legitimate to speak of the Monge-Kantorovich problem in these cases. Moreover, this solution is unique for many important cost functions. All
of these statements can be found in Chapter 1 of [Villani (2003)](#). Unique solutions of the Monge-Kantorovich problem are hence determinable measure preserving maps. The convenience of the Monge-Kantorovich problem lies in the fact that different cost functions $c$ lead to different measure preserving maps and even isomorphisms, many of which occur naturally in Economics and Econometrics. In fact, Optimal Transport theory has found many applications in optimal matching theory (see the applications in [Galichon 2016](#)). The arguably most important measure preserving map which can be derived from the Monge-Kantorovich problem is the Brenier map, which has already seen some application in Statistics ([Carlier, Chernozhukov & Galichon 2016a](#); [Carlier, Chernozhukov & Galichon 2016b](#); [Chernozhukov, Galichon, Henry & Pass 2014](#); [Chernozhukov, Galichon, Hallin & Henry 2016](#)) as a multivariate generalization of strictly increasing and continuous functions. Let us give a brief overview.

### A.1 Brenier map

The Brenier map results from solving the Monge-Kantorovich problem under the standard squared Euclidean distance as a cost function, i.e.

$$
c(y - x) = \|y - x\|^2 = \sum_{i=1}^{d} |y_i - x_i|^2.\$$

[Brenier (1991)](#) first proved that if $P_\varepsilon$ and $P_{Y|X=x}$ possess finite second order moments and if $P_\varepsilon$ is absolutely continuous, then the Monge and the Kantorovich problem coincide, and the unique solution to

$$T_0 := \arg\min_T \int_E \|e - T(e)\|^2 dP_\varepsilon(e), \ T \ \text{measurable} \quad (24)$$

is the gradient of a convex function, i.e. $T_0(e) = \nabla \phi(e)$ for some convex $\phi$. Based on this result, gradients of convex functions are usually called Brenier maps in the Optimal Transport literature. Later, [McCann (1995)](#) proved that the Brenier map always exists and is unique between two probability measures $P_\varepsilon$ and $P_{Y|X=x}$ as soon as $P_\varepsilon$ is absolutely continuous, without the assumption of finite second order moments:

**Theorem 5 (McCann 1995).** Let $P_\varepsilon, P_Y$ be two Borel probability measures on $\mathbb{R}^d$ such that $P_\varepsilon$ vanishes on Borel sets of Hausdorff-dimension $d - 1$. Then there exists a convex function $\phi$ on $\mathbb{R}^d$ whose gradient $\nabla \phi$ is a measure preserving map, pushing forward $P_\varepsilon$ to $P_Y$. Although $\phi$ may not be unique, $\nabla \phi$ is uniquely determined $P_\varepsilon$-almost everywhere.

The assumption that $P_\varepsilon$ vanishes on Borel sets of smaller Hausdorff dimension is actually implied by the assumption that $P_\varepsilon$ be absolutely continuous, so that the theorem in particular holds in the absolutely continuous case. The theorem can also be strengthened in the case where $P_Y$ is also absolutely continuous. Under this additional assumption, there exists $\nabla \phi^*$ such that
for $P_{\epsilon}$ almost every $e$ and $P_Y$ almost every $y$,
\[ \nabla \varphi^* \circ \nabla \varphi(e) = e, \quad \text{and} \quad \nabla \varphi \circ \nabla \varphi^*(y) = y, \] (25)
and $\nabla \varphi^*$ is the $P_Y$ almost everywhere unique gradient of a convex function which pushes forward $P_Y$ to $P_{\epsilon}$ (Villani [2003] p. 67). $\varphi^*$ is the Legendre-Fenchel transform of $\varphi$, see Rockafellar (1997, §12). This can be expressed in a more intuitive way as
\[ T^{-1} = (\nabla \varphi)^{-1} = \nabla \varphi^*, \] (26)
i.e. the Brenier map $T$ is invertible almost everywhere on its domain and its inverse is given by $\nabla \varphi^*$.

Gradients of convex functions are so intriguing for Econometricians, because they define a notion of monotonicity which is often argued to be the most natural generalization of a monotone function to higher dimensions. $T = \nabla \varphi$ is monotone in the following sense (Villani [2003] p. 53):
\[ \langle \nabla \varphi(x) - \nabla \varphi(z), x - z \rangle \geq 0, \]
where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{R}^d$. Here it is easy to see that if $x > z$ in the partial ordering induced by the positive cone on $\mathbb{R}^d$, then this definition implies that $\nabla \varphi(x) \geq \nabla \varphi(z)$ or that $\nabla \varphi(x)$ and $\nabla \varphi(z)$ are not comparable. In particular, $\nabla \varphi$ does not involve crossings in the sense that for $x \in \mathbb{R}^d$
\[ (1 - t)x + t \nabla \varphi(x) = (1 - t)x' + t \nabla \varphi(x') \quad \text{implies that} \quad x = x' \quad \text{for} \quad t \in [0, 1), \] (27)
see the discussion in McCann [1997]; we use this property in the proof of Lemma 3 below. Lastly, note that Theorem 5 also holds for $P_{Y|X=x}$, by disintegration. For a readable introduction to the theory of the Brenier map consider Villani (2003, Chapter 2).

## B Proofs from section 3

### B.1 Proof of Theorem 1

Proof. Since both $g^{-1}(x, z)$ and $h^{-1}(x, z)$ are strictly increasing and continuous, they map, for each $x \in \mathcal{X}_z$, $F_{X|Z=z}(x)$ onto $F_V(g^{-1}(x, z))$ and $F_U(h^{-1}(x, z))$, respectively. Since $F_U(u)$ and $F_V(v)$ are strictly increasing and continuous, it holds that for $u, u' \in \mathcal{U}$ that $F_U(u) = F_U(u')$ if and only if $u = u'$, and analogously for $V$. Now pick a $u \in \mathcal{U}$ with some $F_U(u)$. Since $h$ is strictly increasing and continuous it is measure preserving with measure preserving inverse since both $F_{X|Z=z}$ and $F_U$ are absolutely continuous (Villani [2003] Chapter 2.2). Therefore, for each $z \in \mathcal{Z}$ it holds $F_U(u) = F_{X|Z=z}(h(z, u))$. On the other hand, for each $x \in \mathcal{X}_z$ there exists a unique $v \in \mathcal{V}$ with $F_{X|Z=z}(x) = F_V(g^{-1}(x, z))$ because $g$ is measure preserving with measure
preserving inverse for each \( z \in Z \). Combining these means that for every \( u \in U \) there is a unique \( v \in V \) such that \( v = g^{-1}(h(z, u), z) \) for every \( z \in Z \). Both \( g^{-1} \) and \( h \) are continuous and hence Borel-measurable. The function \( f(u) \equiv g^{-1}(h(z, u), z) \) does not depend on \( z \) by the two previous sentence and is measurable. Moreover, since the problem is perfectly symmetric in \( U \) and \( V \) we can define \( f^{-1}(v) \equiv h^{-1}(g(z, v), z) \) which is also measurable. Therefore, the sigma algebras of \( U \) and \( V \) must coincide. \( \square \)

\[ B.2 \] Proof of Theorem 2

**Proof.** The measurable function \( f : U \to V \) with measurable inverse \( f^{-1} : V \to U \) is fixed by assumption, so that the \( \sigma \)-algebras of \( U \) and \( V \) coincide. Then from Brenier’s multivariate rearrangement results [Brenier (1991)], we know that \( f \) can be uniquely split up as \( f = \nabla \varphi_z \circ s_z \), where \( s_z \) is a measure preserving map transporting \( P_U \) onto \( F_{X|Z=z} \) and \( \nabla \varphi_z \) is the unique gradient of a convex function transporting \( P_{X|Z=z} \) onto \( P_V \) for all \( z \in Z \). \( \nabla \varphi_z \) exists and is unique for all \( z \in Z \) by a theorem in [McCann (1995)] since \( P_V \) and all \( P_{X|Z=z} \) are absolutely continuous. Now define \( g^{-1}(x, z) \equiv \nabla \varphi_z \) and \( h(z, u) \equiv s_z \) for all \( z \in Z \). \( \square \)

\[ B.3 \] Proof of Theorem 3

Here we prove Theorem 3 and its analogue for the case where \( Z \) is absolutely continuous. We need the following important lemmata for the proof.

**Lemma 1.** Let \( F_{X|Z=z} \) and \( F_{X|Z=z'} \) be continuous and strictly increasing with the same support \( \mathcal{X}_z = \mathcal{X}_{z'} \) and let Assumption 9 hold. Let \( T := \nabla \varphi \) be the Brenier map between \( F_{X|Z=z} \) and \( F_{X|Z=z'} \). Then, for each \( x^* \in \mathcal{X}_z = \mathcal{X}_{z'} \), \( T \) is either the metric projection of \( x^* \) onto the epigraph of the isquant \( I_{z'}(x^*) \) or its inverse.

**Proof.** Recall that the graph of \( T, \Gamma \), is cyclical monotone [Gangbo & McCann 1996, Theorem 2.3]. That is, for all \( m \geq 1 \), and for all \( (x_1, Tx_1), \ldots, (x_m, Tx_m) \in \Gamma \),

\[
\sum_{i=1}^{m} \|x_i - Tx_i\|_2^2 \leq \sum_{i=1}^{m} \|x_i - Tx_{i-1}\|_2^2,
\]

with the convention \( x_0 = x_m \).

Now by assumption, the epigraphs \( \text{epi}(F_{X|Z=z}; \alpha) \) and \( \text{epi}(F_{X|Z=z'}; \alpha) \) are closed convex subsets of \( \mathcal{X}_z = \mathcal{X}_{z'} \) for all \( \alpha \in [0, 1] \). Focusing now on some \( \alpha \in [0, 1] \) and a corresponding \( x^* \in \mathcal{X}_z \) such that the corresponding two isquant \( I_z(x^*) = \{ x \in \mathcal{X}_z : F_{X|Z=z}(x) = \alpha \} \) and \( I_{z'}(x^*) = \{ x \in \mathcal{X}_{z'} : F_{X|Z=z'}(x) = \alpha \} \), we know that if they intersect, they do so either by intersecting transversally—which by the transversality theorem yields a closed manifold [Mas-Colell 1989, p. 43]—or coincide, which trivially results in a closed manifold. Also recall that \( T^{-1} := \nabla \varphi^* \) is the Brenier map between \( F_{X|Z=z'} \) and \( F_{X|Z=z} \). With this, our goal is now to show that for each \( x \in I_z(x^*) \) for which \( F_{X|Z=z}(x) > F_{X|Z=z'}(x) \), the Brenier map is the metric projection of \( x \)
onto the closed convex set $\text{epi}(I_{z'}(x^*))$. The metric projection $T$ of $x$ onto the closed convex set $I_{z'}(x^*)$ maps $x$ onto the point $y \in I_{z'}(x^*)$ which is closest to $x$ in the sense that

$$y = \arg\min_{z \in I_{z'}(x^*)} \|x - z\|^2.$$  

This map exists and is unique since $I_{z'}(x^*)$ is a closed convex subset of $\mathcal{X}_z = \mathcal{X}_{z'}$ (Aliprantis & Border 2006, p. 248). By our assumption on the intersection of the manifolds, there are two forms this projection can take: in the case of transversal intersection, it is non-trivial in the sense that the distance between $x$ and its metric projection is positive since $I_z(x^*) \cap I_{z'}(x^*)$ is a manifold by the transversality theorem, and since $x^* \in I_z(x^*) \cap I_{z'}(x^*)$, so that all those $x$ for which $F_{X|Z=z}(x) > F_{X|Z=z'}(x)$ lie outside $\text{epi}(I_{z'}(x^*))$. In the case where the two manifolds coincide the projection is trivial in the sense that the metric projection of $x$ onto $\text{epi}(I_{z'}(x^*))$ coincides with $x$, because $x$ by definition must lie on the boundary of both $I_z(x^*)$ and $I_{z'}(x^*)$. Moreover, since $\mathcal{X}_z = \mathcal{X}_{z'}$, those $x$ are in $\mathcal{X}_z$ just as $\text{epi}(I_{z'}(x^*))$ is, so that there does indeed exist a projection of those $x$ onto $\text{epi}(I_{z'}(x^*))$ in both cases. Now since both $F_{X|Z=z}$ and $F_{X|Z=z'}$ are absolutely continuous, it follows that if a map $T$ is cyclically monotonic, then it must be the Brenier map (Villani 2003, p. 80). Therefore, we only have to show that the metric projection satisfies cyclic monotonicity. So for $m \geq 1$ pick $(x_i, Tx_i)$, $i = 1, \ldots, m$, on $I_z(x^*)$ for which

$$F_{X|Z=z}(x_1) > F_{X|Z=z'}(x_1) \quad \text{and} \quad F_{X|Z=z}(x_2) > F_{X|Z=z'}(x_2),$$

and where $T$ is the metric projection. Then by definition we have

$$\|x_i - Tx_i\|^2 < \|x_i - y_i\|^2 \quad \text{for all} \quad y_i \in \text{epi}(I_{z'}(x^*)) \quad \text{and all} \quad i = 1, \ldots, m,$$

but this immediately implies cyclic monotonicity of $T$ for those points. Another way to see that the Brenier map between $F_{X|Z=z}$ and $F_{X|Z=z'}$ is the metric projection for those $x \in I_z(x^*)$ which lie outside $\text{epi}(I_{z'}(x^*))$ is by noting that the metric projection onto a closed convex set in a Hilbert space is the gradient of a convex function (Moreau 1965, Holmes 1973), and by Theorem 5 this gradient of a convex function is the unique measure preserving map between $F_{X|Z=z}$ and $F_{X|Z=z'}$, i.e. the Brenier map.

Since both $F_{X|Z=z}$ and $F_{X|Z=z'}$ are absolutely continuous, the analogous argument must hold for points $x \in I_{z'}(x^*)$ where $F_{X|Z=z}(x) < F_{X|Z=z'}(x)$, only for $T^{-1}$. Recall that when both distributions are absolutely continuous, the Brenier map is almost everywhere invertible with inverse $T^{-1} = \nabla \varphi^*$, where $\varphi^*$ is the convex conjugate of $\varphi$. Now since $I_z(x^*)$ and $I_{z'}(x^*)$ intersect transversally, just like above it holds that those points $x$ lie outside $\text{epi}(I_z(x^*))$. From the same reasoning it follows that $T^{-1}$ cannot be the Brenier map unless is the metric projection of $x \in I_{z'}(x^*)$ onto the convex set $\text{epi}(I_z(x^*))$. Therefore, the Brenier map $T$ for those $x$ is the inverse of the metric projection onto $\text{epi}(I_z(x^*))$. This shows that $T(I_z(x^*) \cap I_{z'}(x^*)) = I_z(x^*) \cap I_{z'}(x^*)$.\]

55
Lemma 3. Let Assumption 9 hold and consider the Brenier map \( T \) on the disintegration functions \( F_{X|Z=z} \) and \( F_{X|Z=z'} \). If \( 0 < F_{X|Z=z}(x^*) = F_{X|Z=z'}(x^*) < 1 \) for some \( x^* \in X_z = X_z' \), then \( T x^* = x^* \). 

Proof. First note that 

\[
F_{X|Z=z}(x^*) = F_{X|Z=z'}(x^*) \iff P_{X|Z=z}(X \leq x^*) = P_{X|Z=z'}(X \leq x^*),
\]

so that \( T x^* = x^* \) would be measure preserving and hence an admissible solution. The Brenier map in this setting is the metric projection onto the epigraph of the respective isoquant by Lemma 1. But for each point which lies on the intersection between two epigraphs the metric projection is the point itself, so that \( T x^* = x^* \). 

Lemma 4. The first equality in (10) holds. 

Proof. The simplest proof is to consider \( P_{\varepsilon|X=x,Z=z} \) and \( P_{\varepsilon|V=h^{-1}(x,z),Z=z} \) to be disintegrations ([Bogachev 2007, Chapter 10]). Note that these disintegrations exist and coincide with the abstract conditional expectations, because we work in Euclidean space and with absolutely continuous distribution functions ([Chang & Pollard 1997, Theorem 1]). The map \( \phi : (\varepsilon, X, Z) \mapsto (\varepsilon, h^{-1}(X, Z), Z) \) is a measure preserving isomorphism since \( h^{-1} \) is a measure preserving isomorphism for all \( z \in \mathcal{Z} \), just as the identity maps \( \varepsilon \mapsto \varepsilon \) and \( Z \mapsto Z \). But then by Corollary 5.24 in [Einsiedler & Ward 2013] the disintegrated measure \( P_{\varepsilon|V=h^{-1}(x,z),Z=z} \) is the pushforward of the disintegrated measure \( P_{\varepsilon|X=x,Z=z} \), that is 

\[
P_{\varepsilon|V=h^{-1}(x,z),Z=z}(E) = P_{\varepsilon|\phi(x,z)}(E) = P_{\varepsilon|X=x,Z=z}(\phi^{-1}(E)) = P_{\varepsilon|X=x,Z=z}(E),
\]

The following lemma proves the intuitive explanation from page 25 formally.
for all $E \in \mathcal{B}_{\mathbb{R}^d}$. Here, the first equality follows by the definition of $\phi$, the second equality follows by the definition of a pushforward measure as in Corollary 5.24 of [Einsiedler & Ward (2013)], and the third equality follows from the fact that $\phi$ maps $\varepsilon$ to $\varepsilon$.

We can now prove the main theorem of this section.

**Proof of Theorem 3**

Proof. Recall that we want to show identification, so that the assume there is $m : \mathcal{E} \to \mathcal{Y}_x$ as well as $m^* : \mathcal{E} \to \mathcal{Y}_x$ with corresponding unobservable distributions $F_\varepsilon$ and $F_{\varepsilon^*}$ so that $(m, \varepsilon)$ and $(m^*, \varepsilon^*)$ are equivalent in the sense that they generate the same distribution $F_{Y|X=x}$ for endogenous $X$. We work on a suitably equipped complete probability space on which $\varepsilon$, $\varepsilon^*$, $X$, $Z$, and $U$ are defined. Let us first prove that the identified set is

$$I := \{m \in \mathcal{H}(Y_x, \varepsilon) : (m^{-1}(X,Y),U) \perp\!\!\!\!\perp Z\}.$$

The proof of this is almost exactly like the original proof in Torgovitsky (2015a). If $m$ is in the identified set, then $Y = m(X, \varepsilon)$ for some $\varepsilon$ with $(\varepsilon, U) \perp\!\!\!\!\perp Z$. Now since $m$ is a measure preserving isomorphism we have $\varepsilon = m^{-1}(X,Y)$. It therefore holds that $(m^{-1}(X,Y),U) \perp\!\!\!\!\perp Z$, so that $m \in I$. On the other hand, if $m \in I$, then $(m^{-1}(X,Y),U) \perp\!\!\!\!\perp Z$ and since $\varepsilon = m^{-1}(X,Y)$ $m$ is in the identified set. Let us now prove the main part of the theorem.

Since we want to show identification, the goal is to show that $m^* = m$ almost everywhere as in Figure 8, so that $\parallel I$ contains a unique $m$. We can assume that $Z$ is binary with points $z, z' \in Z$ and that $F_{X|Z=z}$ and $F_{X|Z=z'}$ only intersect in one manifold $\mathcal{M}(z, z')$. In fact, if there are more manifolds, it does not matter which one we use; moreover, since the manifolds are of measure zero, their union is still of measure zero, so that we still can identify $q$ modulus the measure zero set of the union of the manifolds.

Now since both $m$ and $m^*$ are isomorphisms and continuous in $x$, $q(x, z, \cdot)$ must be an isomorphism and continuous in $x$ for all $(x, z) \in \mathcal{X} \times \mathcal{Z}$. Recall that by Assumption 6 $q$ does not depend on $z$, so that we can write $q(x, z, \cdot) = q(x, \cdot)$. Furthermore, note that

$$q(x, \varepsilon) = m^{-1}(x, m^*(x, \varepsilon)) \quad \text{for all } (x, z, \varepsilon) \in \mathcal{X} \times \mathcal{Z} \times \mathcal{E}.$$  

Therefore, $q(x, \varepsilon) = \varepsilon$ if and only if $m^* = m$. Now by the assumed normalization, as soon as $q(x, \varepsilon)$ is a function which does not depend on $x$, we know that it must hold that $q(x, \varepsilon) = \varepsilon$, based on the reasoning on page 24, as mentioned, this is analogous to the reasoning in Torgovitsky (2015a), only for measure preserving isomorphisms instead of quantile functions. Therefore, we only need to prove that $q(x, \cdot)$ is independent of $x$.

Now by Assumption 4 $h$ is the identity for $z$ and the Brenier map for $z'$, so that $F_U = F_{X|Z=z}$. Moreover, the fact that $h(\cdot, z)$ and $h(\cdot, z')$ are measure preserving isomorphisms implies
that conditioning on the event \([X = x, Z = z]\) is the same as conditioning on the event \([V = g^{-1}(x, z), Z = z]\) by Lemma 4.

Now in order to prove that \(q\) is constant for almost all \(x\), we use the general sequencing argument mentioned in the main text, similar in spirit to d'Haultfoeuille & Février (2015) and Torgovitsky (2015), only for several dimensions. To make the notation simpler in the following, we denote the Brenier map \(h(z', u)\) mapping \(F_{X|Z=z'}\) onto \(F_{X|Z=z} = F_U\) by \(T\), and its inverse \(h^{-1}(x, z')\) by \(T^{-1}\). We thus want to show that for each \(x \in \mathcal{X}_z = \mathcal{X}_{z'}\) either \(\lim_{n \to \infty} T^n x = x_m\) or \(\lim_{n \to \infty} (T^{-1})^n x = x_m\) for some \(x_m \in \mathcal{M}(z, z')\), so that \(q(x, \cdot)\) is independent of \(x\) up to the measure zero set \(\mathcal{M}(z, z')\). In words, the idea is, for every \(x \in \mathcal{X}_z = \mathcal{X}_{z'}\), to obtain the sequence which does not change \(F_x\), but changes \(F_{X|Z=z}\), so that we can obtain the exogenous effect of \(X\) on the second stage. As stated in the main text, we need to show that this sequence converges, and we will now show that it converges to some element \(x_m \in \mathcal{M}(z, z')\). This will be the difficult step in what follows.

So pick some \(x \in \mathcal{X}_z = \mathcal{X}_{z'}\). If \(F_{X|Z=z}(x) = F_{X|Z=z'}(x)\) there is nothing to prove as by Lemma 2 it holds that \(T^n x = x\) for \(n \in \mathbb{N}\), so that this \(x\) already has converged and must lie in \(\mathcal{M}(z, z')\) since we assume there is only one manifold where the distribution functions intersect. Now depending on the interplay between \(\mathcal{M}(z, z')\) and \(x\) as well as \(F_{X|Z=z}(x)\) and \(F_{X|Z=z'}(x)\), we have to decide if we iterate \(T\) or \(T^{-1}\) to find convergence. The underlying idea is that we always iterate such that the sequence converges towards \(\mathcal{M}\). Now since by part 4 (ii) of Assumption \textcolor{red}{9} all elements \(x\) for which \(F_{X|Z=z}(x) = 0 = F_{X|Z=z'}(x)\) lie on one side of the manifold, which means that for \(x_0\) with \(1 > F_{X|Z=z}(x_0), F_{X|Z=z'}(x_0) > \varepsilon\) for some \(\varepsilon > 0\) it must either hold that \(F_{X|Z=z}(x_0) > F_{X|Z=z'}(x_0)\) or the reverse strict inequality; note that we can rule out the case where equality holds since then this point lies on some manifold and has already converged as just mentioned. So for those points with \(F_{X|Z=z'}(x_0) > F_{X|Z=z}(x_0)\), we iterate \(T\), i.e. from \(F_{X|Z=z'}(x_0)\) to \(F_{X|Z=z}\). The reason is the same as in the one-dimensional case as depicted in the following figure.

![Figure 17: Sequence for a point \(x_0\) in the one-dimensional case](image-url)

\(27\) We denote by \(T^n\) the \(n\)-fold repeated application of \(T\), i.e. \(T^n(x) \equiv T(T(\ldots T(x) \ldots))\).
For the reverse inequality we need to iterate $T^{-1}$ from $F_{X|Z=z}$ to $F_{X|Z=z'}$. Therefore, without loss of generality assume $F_{X|Z=z'}(x) > F_{X|Z=z}(x)$. Then iterating $T$ converges to some $x_m \in \mathcal{M}(z, z')$, because we have

$$F_{X|Z=z'}(T^{n+1}x) \geq F_{X|Z=z}(T^nx)$$

by Lemma 3. Hence, taking the limit $n \to \infty$ by absolute continuity of the two distribution functions yields

$$F_{X|Z=z'}(\lim_{n \to \infty} T^nx) = \lim_{n \to \infty} F_{X|Z=z}(T^nx),$$

which is either a monotonically increasing sequence bounded above by some $x_m \in \mathcal{M}(z, z')$ with $0 < F_{X|Z=z}(x_m) = F_{X|Z=z'}(x_m) < 1$ or a monotonic decreasing sequence bounded below by some $x_m \in \mathcal{M}(z, z')$, depending on which side of $\mathcal{M}(z, z')$ $x$ lies; since those sequences are monotone and bounded in both cases they must converge to $x_m \in \mathcal{M}(z, z')$. This reasoning works in particular, because the range of Brenier maps $T$ and $T^{-1}$ are the respective supports of $F_{X|Z=z}$ and $F_{X|Z=z'}$, so that $T$ never maps outside $\mathcal{X}_z = \mathcal{X}'_z$.

Finally, if there are more manifolds than just $\mathcal{M}(z, z')$, then for some $x$ the respective iterations might converge to different manifolds, but still all elements will converge to some manifold, because all $x$ for which the iterations do not converge to some other manifold, they must converge to $\mathcal{M}(z, z')$ by the above argument. Since all manifolds are of measure zero by assumption, their (finite) union must also be of measure zero.

We are now in the position to conclude the proof by applying the sequencing argument to show that $q(x, e)$ does not depend on almost every $x$. So pick some $x_0 \in \mathcal{X}_z = \mathcal{X}'_z$; we assume that $0 < F_{X|Z=z}(x_0) < F_{X|Z=z'}(x_0)$ for $x_0$ lying on the same side of the manifold as the elements which have zero probability, but the reasoning in the other cases is analogous. By the independence $(\varepsilon, U) \perp Z$ as well as Lemma 4 and the fact that $h$ is a measure preserving isomorphism, we have for every Borel set in $\mathcal{B}_{\mathcal{E}_x}$

$$P_{\varepsilon|X=x_0,Z=z}(E) = P_{\varepsilon|U=h^{-1}(x_0,z),Z=z}(E)$$
$$= P_{\varepsilon|U=h^{-1}(T_{x_0}z'),Z=z}(E)$$
$$= P_{\varepsilon|X=T_{x_0}z,z'}(E) = P_{\varepsilon|X=T_{x_0}z}(E),$$

for every Borel set $E \in \mathcal{B}_{\varepsilon|X}$; analogously,

$$P_{\varepsilon|X=x_0,Z=z}(E') = P_{\varepsilon|U=h^{-1}(x_0,z),Z=z}(E')$$
$$= P_{\varepsilon|U=h^{-1}(T_{x_0}z'),Z=z}(E')$$
$$= P_{\varepsilon|X=T_{x_0}z,z'}(E') = P_{\varepsilon|X=T_{x_0}z}(E').$$

for every Borel set $E' \in \mathcal{B}_{\varepsilon|x_0}$.

Now to conclude the proof of Theorem 5 recall that $q$ is a measure preserving isomorphism mapping each $E' \in \mathcal{B}_{\varepsilon|x_0}$ onto a unique $q(x_0, E')$. Thus as $P_{\varepsilon|X=x_0,Z=z}(E') = P_{\varepsilon|X=T_{x_0}Z}(E')$,
it must hold that \( q(x_0, E') = q(Tx_0, E') \). This holds if we iterate \( T \), so that
\[
q(T^n x_0, E') = \cdots = q(x_0, E') \quad \text{for all} \ E' \in \mathcal{B}_{\varepsilon | X} \text{ and } n \in \mathbb{N},
\]
since \( \mathcal{E}_{x,z} = \mathcal{E} \) by Assumption 6. Now from above we know that the iteration \( T^n x_0 \) converges to some element \( x_m \in \mathcal{M}(z, z') \) and since \( q(\cdot, E') \) is continuous in \( x \), we have
\[
q(x_m, E') = q(\lim_{n \to \infty} T^n x_0, E') = \lim_{n \to \infty} q(T^n x_0, E') = \lim_{n \to \infty} q(x_0, E') = q(x_0, E'). \tag{28}
\]
Therefore \( q(\cdot, E') \) is constant on \( \mathcal{X}_z = \mathcal{X}_{z'} \) modulus \( \mathcal{M}(z, z') \), which by assumption is of measure zero, which concludes the proof.

Lastly, if Assumption 5 holds in place of Assumption 7 then (28) does not hold. However, by definition it holds for all \( \delta > 0 \) that
\[
\lim_{n \to \infty} \mathbb{P}_{\varepsilon | X = x_n, Z = z}(e : |q(T^n x_0, e) - q(x_0, e)| > \delta) = 0 \quad \forall e \in \mathcal{E}_{xz}.
\]
This implies that \( q(x_m, \cdot) = q(x_0, \cdot) \) up to a set of measure zero in \( \mathcal{E}_{xz} \), because it is a well-known result that if a sequence \( \{f_n\}_{n \in \mathbb{N}} \) converges to \( f \) and to \( g \) in measure, then \( f \) coincides with \( g \) almost everywhere. This means that \( q \) is constant on \( \mathcal{X}_z = \mathcal{X}_{z'} \) modulus \( \mathcal{M}(z, z') \) for almost every \( \varepsilon \), so that \( m \) can be identified for almost every \( x \) and almost every \( \varepsilon \).

\[\square\]

### B.4 Identification result for absolutely continuous \( Z \)

In the rather unimportant case where \( Z \) is absolutely continuous we can eliminate Assumption 9 but have to assume that \( m \) is differentiable in \( x \), which is a prohibitively strong assumption in our setting. The statement and proof of this Theorem are analogous to the result in Torgovitsky (2015b), the only difference being that we allow for \( m \) to be a (possibly multivariate) measure preserving \( C^1 \)-diffeomorphism. We still state the whole proof for the convenience of the reader. We denote by \( d_z \) the dimension of \( X \) and by \( d_z \) the dimension of \( Z \).

**Theorem 6** (Identification of nonseparable triangular models with absolutely continuous \( Z \)).

Suppose that \( \mathcal{X} \) is convex, \( Z \) is a continuously distributed vector-valued random variable satisfying Assumption \( 6 \) and \( m(x, z, \varepsilon) : \mathcal{E}_z \to \mathcal{Y}_{zz} \) is a determinable measure preserving isomorphism differentiable in all \( x \) and \( z \) satisfying Assumption \( 7 \). Moreover, let Assumptions \( 3, 5 \) and \( 8 \) hold. Denote by \( G(x, z) \) the \( d_z \times d_x \) matrix with \( (j, k) \) element \( \nabla_{z,j} g^{-1}(x_k, z) \). Then \( m \) is point-identified on \( \mathcal{X} \times \mathcal{E} \) if for every \( x \) in a dense subset \( \mathcal{X}_d \) of \( \mathcal{X} \) and every \( y \in \mathcal{Y}_z \), there exists a \( z \) with \( x \in \mathcal{X}_d \) and \( y \in \mathcal{Y}_{z,z} \) for which \( \nabla_{z} g^{-1}(x, z) \) exists and \( G(x, z) \) exists and has rank \( d_x \).

**Proof.** The proof is analogous to the proof in Torgovitsky (2015b). The ultimate goal is to show that \( q(x, z, \varepsilon) \) from Figure 8 does not depend on \( x \) by differentiation. First, \( q(x, z, \varepsilon) \) is \( C^1 \) since \( m^* \) and \( m \) by assumption are \( C^1 \)-diffeomorphisms. Then analogous to the proof of Theorem 3 we have that \( q(x, z, \varepsilon) = q(x, \varepsilon) \) since \( Z \) is a valid instrument. Now conditioning on the event
[\{U = h^{-1}(x,z), Z = z\}] is the same as conditioning on the event \([X = x, Z = z]\) by Lemma 4. Therefore, by construction we have

\[ q(x, e) = q(x, z, e) = q(u, e) = q(h^{-1}(x,z), e) \]

as depicted in Figure 8. Now fix some \(x \in \mathcal{X}\) and \(e \in \mathcal{E}_{xz}\) so that \(y = m(x,e) \in \mathcal{Y}^0_{xz}\) and take \(\bar{z}\) as in the statement of the theorem. Then \(e \in \mathcal{E}^0_{x,\bar{z}}\) as \(y \in \mathcal{Y}^0_{x,\bar{z}}\) so that we can differentiate \(q(x, z, e)\) in a neighborhood of \((x, \bar{z}, e)\). Now let us differentiate \(q\) first with respect to \(z_j\). This gives by the chain rule

\[ 0 = \nabla_{z_j} q(x, e) = G_j(x, \bar{z}) \nabla_1 q(g^{-1}(x, \bar{z}), e)' \]

where \(G_j(x, \bar{z})\) is the \(j\)-th row of \(G(x, \bar{z})\) and \(\nabla_1\) denotes differentiation with respect to the first argument of \(q(g(v, \bar{z}), e)'\), so that \(\nabla_1 q(g(v, \bar{z}), e)'\) is a \(d_x\) column vector. Stacking the vectors for each \(j\) gives

\[ G(x, \bar{z}) \nabla_1 q(g^{-1}(x, \bar{z}), e)' = 0_{d_x}. \]

By the assumption that \(G(x, \bar{z})\) has full rank, this implies \(\nabla_1 q(g^{-1}(x, \bar{z}), e)' = 0_{d_x}\). Using this information, we can now finalize the proof by differentiating \(q(x, e)\) with respect to \(x_k\) at \((x, \bar{z}, e)\). In fact, since \(\nabla_1 q(g^{-1}(x, \bar{z}), e)' = 0_{d_x}\) this implies by the chain rule that

\[ \nabla_{x_k} q(x, e) = \nabla_{x_k} g^{-1}(x_k, \bar{z}) \nabla_{1,k} q(g^{-1}(x, \bar{z}), e) = 0 \]

for every element \(k\). Here, \(\nabla_{1,k}\) denotes differentiation of the first argument with respect to the \(k\)-th element of \(x\). But this means that \(\nabla_x q(x, e) = 0_{d_x}\) for all \(x \in \mathcal{X}\) as \(\nabla_x q(x, e)\) is continuous and \(\mathcal{X}_d\) is dense in \(\mathcal{X}\) as argued in \cite{Torgovitsky2015b}. By convexity of \(\mathcal{X}\) this implies that \(q(x, e) = r(e)\) for some function \(r\) for every \(x \in \mathcal{X}\) and all \(e \in \mathcal{E}_x\) by continuity. This shows that \(q(x, e)\) does not depend on \(x\) and by the reasoning from the main text this means that \(m\) is identified.

Note that identification here is for all elements in the support, not merely almost all elements.

### C Proof of Theorem 4

The following definition of weak convergence of probability measures is needed, which we have taken from Definition 8.1.1 in \cite{Bogachev2007b}.

**Definition 7.** A sequence of probability measures \(\{P_n\}_{n \in \mathbb{N}}\) on a measurable space \((\mathcal{X}, \mathcal{A}_\mathcal{X})\) is weakly convergent to a probability measure \(P\) if for every bounded and continuous real function \(f\) on \(\mathcal{X}\) one has

\[ \lim_{n \to \infty} \int_{\mathcal{X}} f(x) P_n(dx) = \int_{\mathcal{X}} f(x) P(dx). \]

We denote this convergence by \(P_n \Rightarrow P\).
Moreover, we need the following strengthening of Corollary 5.23 in Villani (2008).

**Lemma 5.** Let $\mathcal{X}$ and $\mathcal{Y}$ be open subsets of $\mathbb{R}^n$ and let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a continuous cost function with $\inf c > -\infty$. Let $\{P_n^X\}$ and $\{P_n^Y\}$ be sequences of probability measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively, such that $P_n^X \leq CP_X$ for all $n \in \mathbb{N}$ and some $0 < C < \infty$. Furthermore, $P_n^X$ converges weakly to $P_X$ and $P_n^Y$ converges weakly to $P_Y$. For each $n$ let $\pi_n$ be the optimal transference plan (i.e. the respective solution of the Kantorovich problem) between $P_n^X$ and $P_n^Y$. Furthermore, assume that

$$\int c\pi_n < \infty \forall n \quad \text{and} \quad \lim_{n \to \infty} \int c\pi_n < \infty.$$  

Suppose also that the corresponding Monge problems have a unique solution and that the optimal transport plans $\pi_n$ and $\pi$ are concentrated on the graph of the optimal transport maps $T_n$ and $T$ solving the respective Monge problems. Then

$$\lim_{n \to \infty} P_n^X(x : |T_n(x) - T(x)| \geq \varepsilon) = 0 \quad \text{for all } \varepsilon > 0.$$  

The proof is verbatim the proof of Proposition 50 in Lindsey & Rubinstein (2016). The idea is to require stronger assumptions on the cost function $c$ to generalize Corollary 5.23 in Villani (2008), which only requires $c$ to be lower semi-continuous, but can only guarantee the result for a fixed $P_X$ and a convergent sequence $P_n^Y$. Since in our case both measures $P_{\varepsilon|X=x,Z=z_i}$ and $P_{\varepsilon|X=x,Z=z_i}$ change with $x$, we need this relaxation which allows both sets of measures to drift.

We are now in the position to prove the important lemma which shows that $m^{-1}(x,y)$ is continuous in measure which we need for our application of Theorem 3.

**Lemma 6.** Under Assumption $T_{14}$, the determinable measure preserving map transporting $P_{\varepsilon|X=x,Z=z_i}$ onto $P_{\varepsilon|X=x,Z=z_i}$ is continuous in measure.

**Proof.** Recall that $F_{X|Z=z}$ and $F_{X|Z=z'}$ are absolutely continuous. Absolutely continuous distributions satisfy Tjur’s property at every point of continuity, which is equivalent to the existence of a unique continuous disintegration $P_{Y|X=x,Z=z} \mapsto P_{Y|X=x,Z=z}$ and analogous for $z'$ in the sense that for each sequence $x_n \to x$, the corresponding probability measures satisfy $P_{Y|X=x_n,Z=z} \Rightarrow P_{Y|X=x,Z=z}$, see Ackerman, Freer & Roy (2016, Lemma 2.7). Moreover, Assumption $T_{14}$ guarantees that $m^{-1}(x,y)$ is a unique and bounded measure preserving map for the Monge-Kantorovich problem under the general cost function $\xi(x,\varepsilon,y)$, which is by definition differentiable and hence continuous. We can hence apply Lemma 5 which guarantees that for every sequence $\{x_n\}_{n \in \mathbb{N}} \in \mathcal{X}_z$ converging to some $x \in \mathcal{X}_z$, the corresponding optimal transport map $m^{-1}(x_n,y)$ converges to $m^{-1}(x,z)$ in probability.  

We can now prove the theorem.
Proof of Theorem 4. Assumption 14 in conjunction with Lemma 6 guarantees that \( m^{-1}(x, y) \) is continuous in measure and that all assumptions are satisfied in order to apply Theorem 3. Therefore, \( m^{-1}(x, y) \) is identified for almost every \( x \) and \( y \). To show identification of \( \bar{U}(x, y) \) we follow the reasoning in Chernozhukov, Galichon, Henry & Pass (2014). Notice that Assumptions 12 and 14 guarantee that \( V(x, y) \) is differentiable. The proof for this is verbatim Step 2 in the proof of Theorem 3(1) in Chernozhukov, Galichon, Henry & Pass (2014). Since \( V(x, y) \) is differentiable and since the inverse demand function \( m^{-1}(x, y) \) is uniquely determined for almost every \( x \) and \( y \), the first order condition

\[
\nabla_y \xi(x, m^{-1}(x, y), y) = \nabla p(y) - \nabla_y \bar{U}(x, y)
\]

identifies \( \nabla_y \bar{U}(x, y) \) for almost every \( x \) and almost every \( y \) so that \( \bar{U}(x, y) \) is almost everywhere identified up to an additive constant, analogous to the reasoning in Chernozhukov, Galichon, Henry & Pass (2014).