Measuring Substitution Patterns in Differentiated Products Industries*

Amit Gandhi[1] and Jean-François Houde[2]

[1]Microsoft & University of Pennsylvania
[2]Cornell University & NBER

December 15, 2017

Introduction

A basic empirical question in industrial organization is the following: which products in a differentiated product market are close competitors with one another. This closeness of competition between two products is determined by the degree of consumer substitutability between them. The measurement of substitution patterns is central to empirical Industrial Organization because it enters many supply side questions of interest. For example, the variation in substitution patterns among the products in a market can be used to study firm "conduct": if there is a high degree of substitutability between the products of rival firms, then markups (and hence prices) should be systematically lower for these products when firms are competing as compared to colluding (Bresnahan 1981, Bresnahan 1987). Furthermore, for any particular hypothesis about firm conduct, substitution patterns drive the effect of counter-factual policy changes on market outcomes, such as mergers, new product introductions, etc.

Although substitution patterns are central to empirical work in imperfectly competitive markets, identifying these substitution patterns from market price and quantity data has proven very challenging.

The mixed-logit model of demand made famous by Berry et al. (1995) (henceforth BLP for short) can in principle approximate very rich substitution patterns by relaxing the strong ex-ante restrictions that the simple logit places on cross price elasticities (most notably the *independence*

*Preliminary and Incomplete. First version: January 26, 2016. Correspondence: Amit Gandhi – Microsoft Research, Seattle, WA, E-mail: agandhi@microsoft.com. Jean-François Houde - Cornell University and NBER, 462 Uris Hall, Ithaca, NY 14850. E-mail: houde@cornell.edu. This research has benefited from the financial support of the NSF (SES-1530788). We thank the many seminar participants who have provided feedback on this paper.

of irrelevant alternatives, aka IIA). This is achieved by allowing consumers to have unobserved taste heterogeneity for observed product characteristics, i.e., random coefficients in utility. While their approach has been hugely influential in providing a framework for studying differentiated product markets, there are very few direct applications (known to us) that have found statistically and/or economically significant departures from the simple logit in practice. The most prominent applications that have successfully recovered non-trivial substitution patterns either use information that is "external" to the mixed logit demand structure, such as supply restrictions (see e.g., Berry et al. (1995), Berry et al. (1999), Eizenberg (2014)), micro moments (see e.g., Petrin (2002), Nielson (2013)), or second choice data (see e.g., Berry et al. (2004), Hastings et al. (2009)), or use more restrictive models of product differentiation such as the nested-logit or GEV models (e.g. Verboven (1996), Bresnahan et al. (1997)).

This user experience has led to a growing questioning of whether consumer heterogeneity in mixed-logit demand systems is even identified with market level data on prices and quantities (see e.g., Metaxoglou and Knittel (2014)). A related challenge for empirical work is that, given the inherent non-linearity of the model, it has been difficult to pinpoint the fundamental variation in the data that drive estimates of substitution patterns in applications. Thus policy conclusions drawn from the model cannot be directly linked to moments in the data that are driving those conclusions (see e.g., Angrist and Pischke (2010)). This has led some to abandon structural demand models altogether in favor of natural experiments to study policy questions in differentiated product markets (see e.g., Ashenfelter et al. (2009)).

In this paper we argue that a potential source for the challenges faced in applied work are weak instruments, and we present a solution to this problem based on a new approximation of the optimal instruments.

Berry and Haile (2014) showed that the parameters governing substitution patters can be identified under fairly general conditions by imposing conditional moment restrictions between the unobserved and observed product characteristics. Letting $\boldsymbol{w}_t = \{\boldsymbol{w}_{1t}, \ldots, \boldsymbol{w}_{Jt,t}\}$ denote the set of exogenous characteristics available in market t, these restrictions are in practice imposed by forming unconditional moment conditions between instruments $A_j(\boldsymbol{w}_t)$, and the unobserved quality of each product computed from the inverse-demand function:

$$E\left[\sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t|\boldsymbol{\theta}) - \boldsymbol{x}_t\boldsymbol{\beta}\middle|\boldsymbol{w}_t\right] = 0 \rightarrow E\left[\left(\sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t|\boldsymbol{\theta}) - \boldsymbol{x}_t\boldsymbol{\beta}\right) \cdot A_j(\boldsymbol{w}_t)\right] = 0, \tag{1}$$

where $\boldsymbol{x}_{jt}\boldsymbol{\beta}$ is the average net quality of product j in market t predicted by the model.

An instrument corresponds to a product specific transformation of the entire menu of observed product characteristics in the market. A first order problem for empirical work is deciding how such instruments *should* be constructed from the data. Our first key point is that the form, $A_j(\cdot)$, of these instruments matters crucially for empirical work - picking arbitrary transformation, or even seemingly intuitive transformations, can readily lead to a weak IV problem which we demonstrate

by way of a simple visual simulation. Weak instruments can explain many of the aforementioned challenges in empirical work. We illustrate this through a series of Monte-Carlo simulations. Using moment conditions commonly used in the literature, we reproduce many of the numerical and estimation problems discussed above, and formally test the weakness of the instruments.

To avoid weak instruments we need to understand how to construct "strong" instruments. The classic results of Amemiya (1977) and Chamberlain (1987) reveal the form of the optimal instruments that achieve the semi-parametric efficiency bound of the model. Although the optimal instruments themselves are infeasible to construct directly from the data, strong instruments that perform well in finite samples will approximate the conditional expectation in equation 1 that defines the reduced-form of the model (as established by e.g., Newey (1993), Ai and Chen (2003), Donald et al. (2003)).

A fundamental problem however for constructing such approximations from the data is that the optimal instruments are a function of *all* the observed product characteristics in a market, and hence even low order approximations to the reduced-form (linear or quadratic) lead to basis function that grow exponentially large in number with the number of products. This can be seen in equation 1: the number of arguments used to define the conditional expectation is proportional to the number of characteristics and the number of products. In typical applications, this number is ofter larger than 1,000, and is proportional to the sample size. This curse of dimensionality has made it impossible to apply insights from the literature on efficient estimation with conditional moment restrictions to the problem of estimating demand for differentiated products.

Our main theoretical contribution is to show that this curse of dimensionality can be solved by using the implicit restrictions that the demand structure of the model places on the reduced-form of the model. In particular we show that the symmetry property implicit in any linear random-utility model implies that the most efficient instruments for a given product/market observation in the data is a *vector symmetric function* of the *differences* between a given product's observed product characteristics and its competitors observed product characteristics. This theoretical result has a powerful Corollary: a finite order approximation to the optimal instrument can be expressed through basis functions that are functions of only these characteristic differences, and importantly the *number* of basis functions is invariant to the number of products in the market. It is this latter result that breaks the curse of dimensionality in approximating the optimal instrument.

Importantly, this characterization of the reduced-form does not depend on the distribution of the random-coefficient, or on the value of the parameters. Therefore, the same instruments can be used to estimate different models of product differentiation, and to test between alternative specifications (e.g. correlated random-coefficients, different distributional assumptions, etc.). This is an important advantage of our approach over alternative two-step approximations to the optimal instruments that require obtaining consistent estimates of the parameters, and rely on a correctly specified model to be valid.

The basis functions themselves are economically meaningful – they capture the relative isolation of each product in characteristic space – and have not yet been systematically exploited in empirical work. If we apply these basis functions directly as instruments in a GMM estimation we can approximate (arbitrarily well) the performance of the optimal instruments (following Theorem 5.2 in Donald et al. (2008)).[1] Our Monte-Carlo simulation results show that exogenous measures of product differentiation, or Differentiation IVs, lead to substantial improvements in the small sample performance of the GMM estimator. In particular, we illustrate that simple forms of those instruments can eliminate the weak identification problem, and identify rich models of substitution patterns with large numbers of random-coefficients and correlated taste shocks.

Related Literature

Our IV's are a natural complement to the large literature on price instruments in differentiated product markets. Price endogeneity is a familiar problem in the literature with a long history, and a variety of instruments have now been proposed to address it, i.e, BLP instruments, Haussman instruments, Waldfogel instruments, etc.[2] However, a key point in Berry and Haile (2014) is that the identification of substitution patterns poses a distinct empirical problem from price endogeneity.[3] This is because there are in fact *two* different sets of endogenous variables in the model - prices *and* market shares - which require different sources of exogenous variation for the model to be identified. However the literature has been virtually silent about the appropriate form of the instruments for market share? We believe the root of the problems encountered in empirical practice is that there does not exist any formal discussion of how to construct such instruments, and thus researchers have used a single set of instruments, namely price instruments - i.e., instruments constructed on the basis of what should vary price in the model - as instruments for both prices *and* markets shares. Our approximation is based on first isolating the endogeneity problem induced by market shares, which is the key that allows us to derive the vector symmetric form of the optimal instrument. Our approach however naturally extends to allowing for endogenous prices. This extension combines our approximation with the Berry et al. (1999) approach to optimal IV that was revisited recently by Reynaert and Verboven (2013), and shows that these alternative perspectives

[1] An alternative approach to deal with weak instruments is to estimate the model using estimators that are robust to weak identification (e.g. Stock and Wright (2000)). Conlon (2013) for instance describes the properties of an Empirical Likelihood-based estimator applied to BLP, and demonstrates a weak identification problem associated with commonly used instruments.

[2] Price endogeneity is linked directly to the classic simultaneous equations problem of prices and quantities being simultaneously determined in market equilibrium and is common to both homogenous good and differentiated product markets. A natural instrument for prices is to use a cost side instrument, but such cost instruments are often not immediately available. The well known "BLP instruments" provide an alternative source for variation in prices in differentiated product settings that is based on a first order approximation of the equilibrium pricing function. BLP IV's comprise of sums of product characteristics of competing products interacted with ownership structure, and are the standard instruments used in mixed logit demand applications.

[3] Although they consider a non-parametric form of the model, this conclusion applies with equal force to the standard parametric specification used in practice.

are indeed complementary to one another in an empirically powerful way. Our approach also extends to settings in which the distribution of demographic characteristics vary across markets (similar to Romeo (2010)). Finally, the instruments that we propose are similar to instruments commonly used to identify nested-logit and spatial differentiation models.[4] A key contribution of our paper is to proposed a unifying approach to study the identification of random-coefficient models with aggregate data.

The rest of the paper proceeds as follows. We beginning by illustrating the problems associated with weak identification in the context of a mixed-logit model with exogenous characteristics. In Section 2 we illustrate how symmetry of the demand can be used to solve the curse of dimensionality problem in approximating the reduced-form of the model. We then use the symmetry property to characterize relevant instruments that can be used to identify the parameters determining substitution patterns, i.e. Differentiation IVs. Finally, in Section 3 we illustrate the small sample performance of those instruments in different contexts, including settings with many random-coefficients, correlated tastes, and endogenous characteristics. We also contrast our approach with the related approximation to the optimal IVs proposed by Berry et al. (1999) and Reynaert and Verboven (2013).

1 Baseline model: Exogenous characteristics

In order to illustrate the instrument choice problem, we consider a special case of the random-utility model considered by Berry, Levinsohn, and Pakes (1995) [BLP], in which product characteristics (including prices) are exogenous. We first describe the model notation that we use throughout the rest of the paper, and discuss the identification problem. We then illustrate the problems associate weak instruments through a simulation example.

1.1 Notation and assumptions

Consider a market t with $J_t + 1$ differentiated products. Each product j is characterized by a vector of observed (to the econometrician) product characteristics $\boldsymbol{x}_{jt} \in \mathbb{R}^K$ and an unobserved characteristic ξ_{jt}. We will refer to $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{J_t,t})$ as a summary of the observed *market structure* - the entire menu of observed product characteristics available to consumers in market t (i.e. $J_t \times K$ matrix). Similarly, $\boldsymbol{s}_t = \{s_{1t}, \ldots, s_{J_t,t}\}$ is the vector observed market shares, which is defined such that $1 - \sum_{j=1}^{J_t} s_{jt} = s_{0t}$ is the market share of the "outside" good available to all consumers in market t. We normalize the characteristics of the outside good such that $\boldsymbol{x}_{0t} = 0$ and $\xi_{0t} = 0$.[5]

[4] See in particular Berry (1994), Bresnahan et al. (1997), Pinkse et al. (2002), Davis (2006), Thomadsen (2007), and Houde (2012).

[5] Thus each characteristic can be interpreted in terms of differences relative to the outside good.

We consider initially a model in which the unobserved quality ξ_{jt} is independent of the characteristic of products available in market t. This identifying assumption takes the form of the conditional mean restriction (CMR) defined in Assumption 1.

Assumption 1. *The unobserved quality of products has mean zero conditional on the observed menu of characteristics* $\boldsymbol{x}_t$*,*

$$E\left[\xi_{jt} \mid \boldsymbol{x}_t\right] = 0. \tag{2}$$

Following, Berry and Haile (2014), the product characteristics space is partitioned into two subsets: $\boldsymbol{x}_{jt}^{(1)}$ refers to *linear* product characteristics, and $\boldsymbol{x}_{jt}^{(2)}$ to *non-linear* product characteristics. The economic distinction between the two groups of attributes is that consumers are assumed to have *homogenous* tastes for linear characteristics. Let K_2 denotes the dimension of the $\boldsymbol{x}_{jt}^{(2)}$ vector.

We assume that consumers have linear preferences for product characteristics:

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^{K_2} \nu_{ik} x_{jt,k}^{(2)} + \epsilon_{ijt} \tag{3}$$

where $\delta_{jt} = \boldsymbol{x}_{jt}'\boldsymbol{\beta} + \xi_{jt}$ is labelled the "mean utility" of product j, and $\epsilon_{ijt} \sim \text{T1EV}(0,1)$ is the an idiosyncratic utility shock. According to this specification, the taste of consumer i for characteristic k belonging to $\boldsymbol{x}_{jt}^{(2)}$ is given by $\beta_k + \nu_{ik}$. The random-coefficients vector, $\boldsymbol{\nu}_i$, is distributed according to a density function $\phi(\cdot;\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of non-linear parameters to be estimated.[6]

If each consumer i chooses the product that maximizes his/er utility, we can integrate over the distribution of consumer random utilities to construct the demand function for product j:

$$\sigma_j\left(\boldsymbol{\delta}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}\right) = \int \frac{\exp\left(\sum_k v_{ik} x_{jt,k}^{(2)} + \delta_{jt}\right)}{1 + \sum_{j'=1}^{n_t} \exp\left(\sum_k v_{ik} x_{j't,k}^{(2)} + \delta_{j't}\right)} \phi(\boldsymbol{\nu}_i;\boldsymbol{\lambda}) d\boldsymbol{\nu}_i \tag{4}$$

where $\boldsymbol{x}_t^{(2)} = \left(\boldsymbol{x}_{1t}^{(2)}, \ldots, \boldsymbol{x}_{n_t,t}^{(2)}\right)$ and $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{n_t,t})$.

Following Berry (1994), the *inverse demand* function is used to define the residual function of the model:

$$\begin{aligned} s_{jt} &= \sigma_j\left(\boldsymbol{x}_t^{(2)}, \boldsymbol{\delta}_t; \boldsymbol{\lambda}\right) \quad j = 1, \ldots, J_t \\ &\iff \rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) = \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}\right) - \boldsymbol{x}_{jt}\boldsymbol{\beta} \quad j = 1, \ldots, J_t \end{aligned} \tag{5}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$ is the full parameter vector of dimension m. Existence and uniqueness of the inverse demand, $\sigma_j^{-1}(\cdot)$, follows directly from Berry (1994), Berry, Levinsohn, and Pakes (1995). See also Berry, Gandhi, and Haile (2013) for a general proof that does not rely on the type-1

[6] Note that this density is assumed to be common across market. We relax this assumption below when studying the role of demographic characteristics variation across markets.

extreme-value distribution assumption of ϵ_{ijt}.

1.2 Identification and estimation

Equation (5) highlights the need for instruments. Intuitively, since consumers observe ξ_{jt} before choosing which product to buy, the vector of observed market shares is correlated with the residual at true value of the parameters $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. As discussed first by Jorgensen and Laffont (1974) and Amemiya (1974), this implies that the non-linear least-square estimator (NLSE) of $\boldsymbol{\lambda}$ suffers from a standard simultaneity bias problem (even when all x's are exogenous). To see this, note that the first-order conditions of NLSE with respect to $\boldsymbol{\lambda}$ are not satisfied at the true value of the parameters:

$$E\left[\frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^0)}{\partial \boldsymbol{\lambda}} \cdot \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0)\right] = E\left[\frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^0)}{\partial \boldsymbol{\lambda}} \cdot \xi_{jt}\right] \neq 0.$$

In the reminder of the paper, we refer to the inverse demand, $\sigma_j^{-1}(\cdot)$, as the *structural* equation of the model.

To get around this problem, Berry (1994) and Berry, Levinsohn, and Pakes (1995) propose to estimate $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ by GMM. Formally, one need to construct a vector $\boldsymbol{z}_{jt}$ of exogenous instruments from $\boldsymbol{x}_t$, which, consistently with the restrictions imposed by equation 2 above, satisfy the following $L > m$ unconditional moment restrictions:

$$E\left[\rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right) \cdot \boldsymbol{z}_{jt}\right] = 0. \tag{6}$$

where $\boldsymbol{\theta}^0$ denotes the $m \times 1$ vector of parameters.

The non-linear GMM problem is defined as follows:

$$\min_{\boldsymbol{\theta}} \quad n g_n(\boldsymbol{\theta}) \boldsymbol{W}_n g_n(\boldsymbol{\theta})^T \tag{7}$$

where $\boldsymbol{W}_n$ is an $L \times L$ efficient weighing matrix, and $g_n(\boldsymbol{\theta}) = \boldsymbol{\rho}(\boldsymbol{\theta})^T \boldsymbol{Z}/n$ is the empirical counterpart of the moment conditions defined in equation (6).

Since the GMM objective function is a quadratic form, the Gauss-Newton (GN) algorithm is a computationally efficient method for finding the minimum. Each optimization step is obtained by estimating a linear GMM problem corresponding to a linear approximation of the residual function.

Algorithm 1. *Initiate the algorithm at parameter* $\boldsymbol{\theta}^1$*. Iteration* k*:*

1. *Invert demand system at* $\boldsymbol{\theta}^k$*:* $\rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^k\right) = \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^k\right) - \boldsymbol{x}_{jt}\boldsymbol{\beta}^k$

2. Evaluate the Jacobian of the residual-function using the implicit function theorem:

$$\frac{\partial \rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^k\right)}{\partial \boldsymbol{\theta}^T} = \left\{ -\boldsymbol{x}_{jt}, \frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^k\right)}{\partial \boldsymbol{\lambda}^T} \right\} = \boldsymbol{Y}_{jt}(\boldsymbol{\theta}^k)$$

3. Compute the Guass-Newton step using linear GMM:

$$\rho_{jt}(\boldsymbol{\theta}^k) = \boldsymbol{Y}_{jt}(\boldsymbol{\theta}^k)\boldsymbol{b} + e_{jt} \Rightarrow \hat{\boldsymbol{b}} = \left((\boldsymbol{Y}^T\boldsymbol{Z})\boldsymbol{W}_n(\boldsymbol{Z}^T\boldsymbol{Y})\right)^{-1}(\boldsymbol{Y}^T\boldsymbol{Z})\boldsymbol{W}_n(\boldsymbol{Z}^T\boldsymbol{\rho})$$

4. Update parameter vector:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \hat{\boldsymbol{b}}$$

5. Stop if $||\hat{\boldsymbol{b}}|| < \varepsilon$. Else repeat steps 1-5.

The Gauss-Newton algorithm has good convergence properties when the moments are strong. This is because strong instruments imply a lot of curvature in the GMM objective function, which is therefore well approximated by a quadratic function. In contrast, weak instruments are associated with little or no curvature in the objective function, which leads to convergence problems (see below for numerical examples).

The Gauss-Newton algorithm also highlights the fact the model can be represented by a linear GMM problem. Step (3) corresponds to a Gauss-Newton regression. The solution, $\hat{\boldsymbol{\theta}}$, is implicitly defined by setting the linear parameters of Gauss-Newton regression to zero: $\hat{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) = 0$. This defines a linear reduced-form for the GMM problem:

$$\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}) = \boldsymbol{Z}\boldsymbol{\pi}\boldsymbol{b} + \boldsymbol{v}^1 \tag{8}$$

$$\boldsymbol{J}(\hat{\boldsymbol{\theta}}) = \boldsymbol{Z}\boldsymbol{\pi} + \boldsymbol{v}^2 \tag{9}$$

where $\boldsymbol{J}(\hat{\boldsymbol{\theta}})$ is a $n \times |\boldsymbol{\lambda}|$ matrix containing the slopes of the inverse demand with respect to each of the non-linear parameters (i.e. $J_{jt,k}(\boldsymbol{\theta}) = \partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^k\right)/\partial \lambda_k$), $\boldsymbol{\pi}$ is a $K \times |\boldsymbol{\lambda}|$ matrix of reduced-form parameters, and $(\boldsymbol{v}^1, \boldsymbol{v}^2)$ are the reduced-form residuals. Standard rank conditions for local identification of the model requires that the moment conditions contain enough excluded instruments correlated with the slope of the inverse demand (i.e. the endogenous variables of the model).

Berry and Haile (2014) formally showed that the demand function can be non-parametrically identified by assuming that the unobserved product attributes are conditionally independent of the menu of product characteristics. To see this in our context, take the conditional expectation of the residual function defined in equation (5):

$$E\left[\rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0)|\boldsymbol{x}_t\right] = E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^0\right)\middle|\boldsymbol{x}_t\right] - \boldsymbol{x}_{jt}\boldsymbol{\beta}^0 = E\left[\xi_{jt}|\boldsymbol{x}_t\right] = 0, \tag{10}$$

where the expectation operator is taken over the J_t endogenous variables: $\{s_{1t}, \ldots, s_{J_t,t}\}$. We will refer to the conditional expectation of the inverse-demand function as the *reduced-form* of the model.

The logic of the identification argument relies on an exclusion restriction: consumers have homogenous tastes for at least one continuous exogenous characteristic (i.e. *special regressor*). Since all characteristics included in $\boldsymbol{x}_{jt}^{(1)}$ are excluded from the structural equation $\sigma_j^{-1}(\cdot)$ due to the linearity of the utility function, the reduced-form of the model includes more exogenous variables than the structural equation (i.e. inverse-demand).

Intuitively, the *linear* characteristics of rival products can be used as "instruments" for market shares to non-parametrically identify $\sigma_j^{-1}(\cdot)$. This result is important for the identification of parametric model as well, since it clearly indicates that the conditional mean restriction implies a large number of over-identifying restrictions. The challenge for empirical work however is how to construct *relevant* instruments that are consistent with those restrictions.

1.3 Illustration of the weak identification problem

The problems of weak identification in linear IV models are well documented (e.g. Stock, Wright, and Yogo (2002)). The weak correlation between the endogenous variables and the excluded instruments leads to biased estimates, and a failure of the standard asymptotic distribution approximation.

The same problems arise in non-linear IV models, but are in general harder to diagnose. To see this, recall that the model is identified if the following two conditions are satisfied:

$$E\left[\rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) \cdot \boldsymbol{z}_{jt}\right] = 0 \text{ iif } \boldsymbol{\theta} = \boldsymbol{\theta}^0 \tag{11}$$

$$\operatorname{rank}\left(E\left[\frac{\partial \rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^T} \cdot \boldsymbol{z}_{jt}\right]\right) = m \tag{12}$$

The first condition refers to the global identification of model, while the second equation is concerned with local identification.

In the context of the demand model, the two conditions ensure that the instrument variables $\boldsymbol{z}_{jt}$ are correlated with the inverse demand *away* from the true parameters values. Intuitively, the model is identified if the IVs are able explain the assignment of the products' unobserved qualities when the inverse demand mapping is evaluated at the "wrong" parameter values. In contrast, the instruments are weak if the implied inverse demand is (almost) orthogonal to the instrument vector when evaluated away from the true parameter values.

To fix ideas, consider the following single dimension example:[7]

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + (\beta_2 + \lambda\eta_i) \cdot x_{jt}^{(2)} + \xi_{jt} + \epsilon_{ijt}, \quad j = 1, \ldots, 15 \text{ and } t = 1, \ldots, 100$$

where $\eta_i \sim N(0,1)$ and $\epsilon_{ijt} \sim \text{T1EV}(0,1)$. From this example, we construct the following instrumental variable:

$$\text{IV}_{jt}^{\text{sum}} \quad = \quad \sum_{j' \neq j}^{15} x_{j',t}^{(2)}$$

The sum of rival characteristics is a commonly used instrument in the literature (see Berry, Levinsohn, and Pakes (1995)).

To illustrate the weakness of this instrument with respect to λ, we beginning by plotting the joint distribution of the instrument and the error evaluated at the "wrong" parameter value. We focus in particular on the inverse demand function evaluated under multinomial Logit preferences: $\sigma_j^{-1}(\boldsymbol{s}_t; \lambda = 0) = \log(s_{jt}) - \log(s_{0t})$ (Berry (1994)).

Figure 1a plots this residual quality, against the "residualized" instrument.[8] Each dot represents a product/market combination, and the line corresponds to a linear regression of $\rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta})$ on the instrument. As the figure illustrates the sum of rival characteristics is uncorrelated with the inverse demand evaluated at $\lambda = 0$. The R^2 and the slope of the regression are both indistinguishable from zero. In other words, the moment conditions are (nearly) satisfied away from the true parameter value ($\lambda^0 = 4$), implying that the moment restriction weakly identify the model.

Importantly, this weak identification problem is not caused by a small sample problem ($N = 1500$), or a lack of variation across products (as discussed in Armstrong (2014)). To illustrate both points Figure 1b uses the same data-set to measure the correlation between the estimated residual under multinomial logit preferences and the *Euclidian* distance between products.[9] Unlike with the previous instrument, this variable is strongly correlated with the model residual evaluated at $\lambda = 0$; the R^2 of the regression after partialling-out the effect of $\boldsymbol{x}_{jt}$ is over 0.35 (compared to 0.0006 in Figure 1a).

The sign of the correlation between distance and $\sigma_j^{-1}(\boldsymbol{s}_t; \lambda = 0)$ is also intuitive. The inverse demand evaluated at $\lambda = 0$ is a (log) linear transformation of product market shares. Since the data is generated by a model with non-IIA preferences, products located in denser areas of the product space have relatively small market shares. The inverse demand evaluated at $\lambda = 0$ rationalizes this feature by assigning high unobserved quality to products that are relatively isolated,

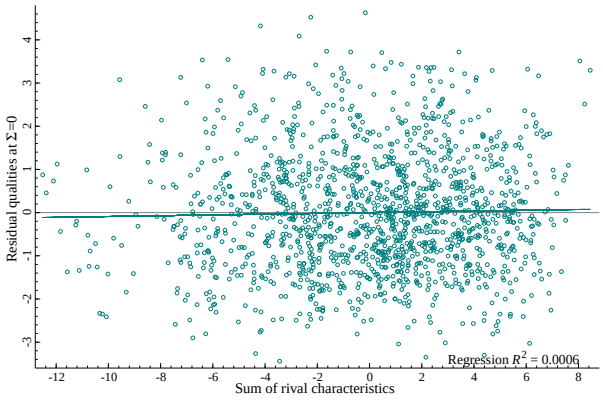
[7] The data-set for this example is generated as follows: $x_{jt}^k \sim N(0,1)$ for $k = 1, 2$ and $\xi_{jt} \sim N(0,1)$. The parameter values are given by: $\beta_0 = -3, \beta_1 = 1, \beta_2 = 1, \lambda = 4$.

[8] The x-axis was obtained by first regressing the excluded instrument on an intercept and the two continuous characteristics.

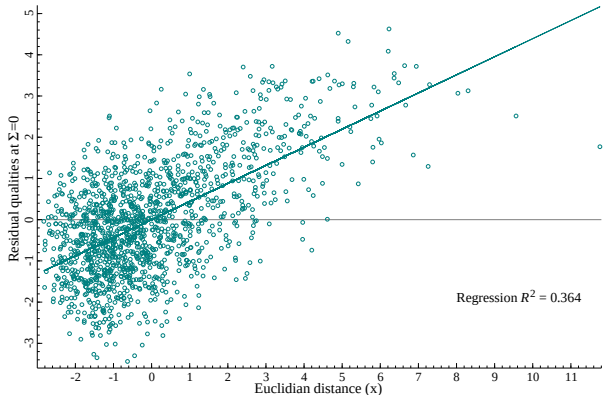
[9] The instrument is constructed as follows: $\text{IV}_{jt}^{\text{dist}} = \sqrt{\sum_{j' \neq j}^{15} \left(x_{j',t}^{(2)} - x_{jt}^{(2)} \right)^2}$.

Figure 1: Scatter plot of residual product qualities under multinomial Logit preferences against weak and strong instrumental variables

(a) Weak Instrument



(b) Strong Instrument



and low residual quality to products with many substitutes. This positive relationship between differentiation (or distance) and the inverse demand at $\lambda = 0$ is captured by the strong instrument used in Figure 1b, but not by the weak instrument in Figure 1a.

To formalize this intuition, we extend the previous example by adding multiple dimensions of unobserved heterogeneity:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + \sum_{k=1}^{K_2} (\beta_{2,k} + \lambda_k \eta_{ik}) \cdot x_{jt,k}^{(2)} + \xi_{jt} + \epsilon_{ijt}, \quad j = 1, \ldots, 15 \text{ and } t = 1, \ldots, 100. \quad (13)$$

Using the previous notation, $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_{K_2}\}$ denotes the vector of K_2 non-linear parameters. For each specification, we use the sum of characteristics of rivals' products as instruments (including $x_{jt}^{(1)}$), which leads to $K_2 + 1$ IVs and one exclusion restriction (i.e. $\boldsymbol{z}_{jt} = \{1, x_{jt}^{(1)}, \boldsymbol{x}_{jt}^{(2)}, \text{IV}_{jt}\}$).[10] As before, the characteristics of products are generated assuming standard-normal distributions, and the structural residuals have equal variances across products and markets.

The results of 1,000 Monte-Carlo replications are summarized in Table 1. The first four rows report the average bias and root-mean square error (RMSE) of the estimated log parameters. Note that we estimate the log of σ_k in equation (13), instead of σ_k directly, to account for the strictly positive support of the parameter space. The next four rows report the bias and RMSE of the transformed estimated parameters. We use the results of the simulations to document three important features associated with weak instruments: (i) failure to reject the full rank and IIA hypothesis, (ii) imprecise estimates, (iii) optimization problems.

[10] The number of exclusion restrictions in this example corresponds to the number of "linear" characteristics: $|x_{jt}^{(1)}| = 1$.

Table 1: Monte-Carlo simulation results for exogenous characteristics model with weak instruments

	$K_2 = 1$		$K_2 = 2$		$K_2 = 3$		$K_2 = 4$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\log \lambda_1$	-11.293	95.930	-5.433	74.954	-1.147	5.503	-8.400	229.670
$\log \lambda_2$			-4.692	58.306	-1.364	6.261	-1.096	6.173
$\log \lambda_3$					-1.407	9.199	-4.657	112.637
$\log \lambda_4$							-0.926	4.023
λ_1	0.136	2.643	-0.010	2.486	-0.032	2.195	0.218	2.348
λ_2			0.117	2.421	-0.006	2.267	0.099	2.297
λ_3					0.178	2.377	0.113	2.378
λ_4							0.075	2.207
1(Local-min)	0.189		0.514		0.594		0.661	
Range(J-statistic)	0.737		1.149		1.636		1.513	
Range(p-value)	0.167		0.189		0.212		0.210	
Range(param)	11.735		6.641		6.583		4.863	
Rank-test	1.265		0.464		0.259		0.178	
p-value	0.615		0.813		0.886		0.919	
IIA-test	1.327		1.296		1.486		1.944	
p-value	0.426		0.422		0.356		0.237	

Feature 1: Weak identification tests

We propose two tests to measure the weakness of the instruments.

The first one, *Rank-test*, is testing the null hypothesis that the matrix $E\left[\partial \rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) / \partial \boldsymbol{\theta}^T \cdot \boldsymbol{z}_{jt}\right]$ has rank m (i.e. local identification assumption). In non-linear IV models, testing the rank of this matrix is not always feasible since the Jacobian matrix depends on the unknown parameter vector $\boldsymbol{\theta}$ (see Wright (2003) for a discussion). This is not an issue in our Monte-Carlo experiment, since we can test the hypothesis at the true value of parameter. In this case, the rank-test proposed by Cragg and Donald (1993) has the correct size asymptotically to test the null hypothesis of under-identification. This test can easily be implemented using standard statistical softwares by estimating the reduced-form of the model, defined in equation (**??**), evaluated at $\boldsymbol{\theta}^0$ (e.g. *ranktest* or *ivreg2* in STATA). The results confirm that the sum of rival characteristics are very weak instruments in our context. The null hypothesis of under-identification (i.e. rank less than m) cannot be rejected with probability ranging between 60% and 92% on average across the specifications.

Our second approach is an *ex-ante* specification test aimed at evaluating the ability of the instruments to reject alternative models (i.e. global identification). In line with the discussion of Figure 1a it is feasible in the context of the mixed-logit model to test the validity of a particular "wrong" model: the multinomial logit model.

Hausman and McFadden (1984) proposed a simple specification test of the IIA hypothesis with

micro-data: estimate the model by including characteristics of rival products in the indirect utility of consumers, and test the exclusion restriction implied by the multinomial logit model. This insight can be used in our context by testing the validity of the exclusion restrictions at $\boldsymbol{\Sigma} = 0$. For instance, in our exogenous characteristic example, this amounts to test the restriction that the instrument vector, which measures the characteristics of rival products, should be excluded from the average utility of product j (relative to 0). Since the inverse-demand under the null is linear, this hypothesis can be tested as follows:

$$\rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\beta}, \boldsymbol{\lambda} = 0\right) = \ln s_{jt}/s_{0t} = \boldsymbol{x}_{jt}\boldsymbol{\beta} + \boldsymbol{z}_{jt}\boldsymbol{\gamma} + \xi_{jt}. \tag{14}$$

If Assumption 1 is valid, the IIA property corresponds to the null hypothesis that $\hat{\boldsymbol{\gamma}} = 0$. This test can be constructed using standard linear regressions techniques (i.e. avoiding any non-linear optimization).[11]

If the true model is such that $\boldsymbol{\lambda}^0 \neq 0$, failing to reject the IIA hypothesis is a sign that the moment conditions are weak. Obviously, $\boldsymbol{\lambda} = 0$ is only one particular potentially "wrong" model. However since it is a computationally simple test to conduct, failure to reject the Logit model is a strong signal that alternative instruments should be considered, and/or that the Logit model might be the best model to use given the data.[12]

The last two lines of Table 1 report the average F-statistics and p-values associated with the null-hypothesis of IIA preferences (i.e. $\hat{\boldsymbol{\gamma}} = 0$). As with the Cragg-Donald statistics, we cannot reject the null hypothesis that the model is weakly identified. In particular, across all four specifications, the instruments cannot reject IIA hypothesis using standard confidence levels.

Feature 2: Bias and precision of the estimates

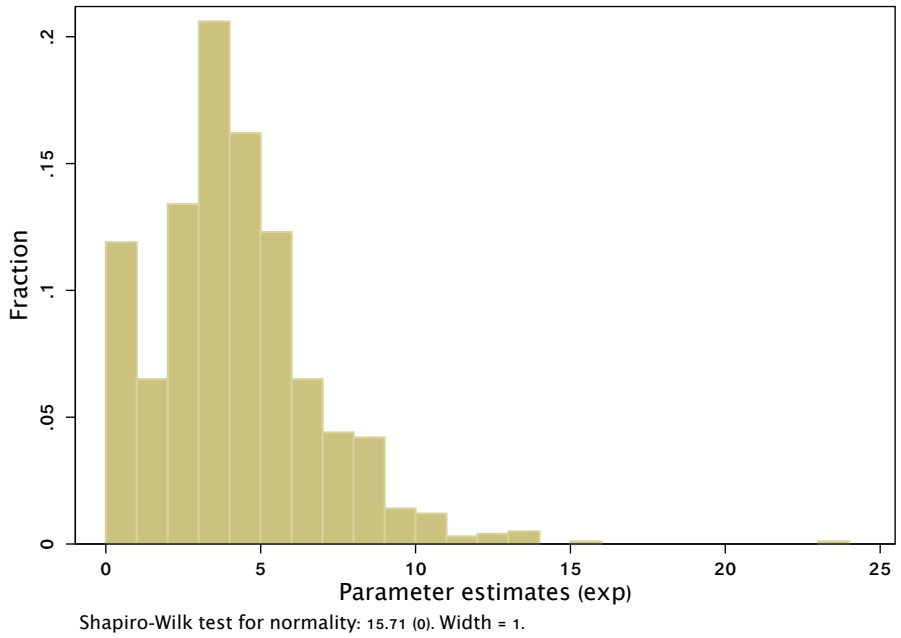
The second striking feature of the simulation results are the large negative biases and important dispersion in the estimates of the log parameters. This poor performance is largely explained by the presence of large negative outliers causing the estimated standard-deviation to be approximately zero (i.e. $\hat{\sigma}_k \approx 0$). Figure 2 plots the distribution of $\hat{\sigma}_1$ in the $K_2 = 1$ specification. In this specification, 8.4% of $\hat{\sigma}_1$ are estimated to be less than 0.001, which can be interpreted as a corner solution to the GMM optimization problem. This is a robust feature of weak instruments that has been documented by other researchers analyzing the BLP model (e.g. Reynaert and Verboven (2013)).

It is caused by a combination of two factors. First, weak instruments imply that the normal

[11] When $\boldsymbol{x}_{jt}$ includes endogenous variables (e.g. prices), the same IIA hypothesis can be tested by testing the validity of the over-identification restrictions, or by using separate price instruments. See discussion below.

[12] Obviously, the IIA hypothesis can be rejected because the instruments are not valid, and not because the instruments are strong. If that is the case, the over-identifying restrictions should also be rejected with the non-linear model. It is therefore important that researchers report both tests (i.e. IIA-test and over-identifying restriction test) when presenting their results.

Figure 2: Distribution of the random-coefficient parameter estimate with weak instruments



distribution is a poor approximation of the finite-sample distribution of the parameter estimates. This can clearly be seen in Figure 2; we can easily reject the null hypothesis of normality using Shapiro-Wilk test statistic. In practice, this means that the distribution of the parameter estimates produce frequent outliers. Second, since the parameter space is bounded, outliers on the left-side of the distribution lead to corner solutions. This has important implications for applied work. Not only does it complicate inference, but it also lead to frequent rejection of models with heterogenous taste for product attributes.

The second panel in Table 1 shows the dispersion of parameter estimates is still large after transforming the parameter estimates. The RMSEs range from 2.2 to 2.6 across specifications; or more than 50% of the true parameter value (i.e. $\sigma_k = 4$ for all k's). The precision of the estimates is poor across all four specifications, and does not deteriorate when we increase the number dimensions of heterogeneity. Notice also that the sample sizes are fairly large relative to the number of parameters to estimate (i.e. $1,500$ versus $K_2 \leq 4$). Therefore, as with linear IV models, the results confirm that weak instruments lead to highly imprecise estimates.

Feature 3: Numerical optimization problems

Another consequence of weak instruments in non-linear IV problems is the presence of local minima and numerical optimization problems. To illustrate this point, for each simulated sample, we launch the optimization routine at 10 random starting values (centered around the true), and use a Nelder-

Mead (or Simplex) algorithm to find the local minimum. The indicator variable 1(Local-min) is equal to one if the algorithm converged to more than one solution.[13] This procedure clearly shows that multiple local minima is a frequent phenomenon in our simulations with weak instruments. Moreover, the frequency of the problem is increasing with the dimensionality of the parameter space. When $K_2 = 4$, 66% of the samples exhibit multiple minima out of 10 starting values, compared to 19% when $K_2 = 2$.

The next two rows illustrate the magnitude of the differences between the different local solutions. For the samples exhibiting multiple solutions, *Range(J-stat)* and *Range(p-value)* calculate the average difference between of largest and smallest J-statistics and p-values respectively, while *Range(param)* calculates the average absolute difference between the parameter estimates. The average differences in the J-statistic p-values imply that the over-identifying restrictions are rejected with a p-value of roughly 20% on average using the largest local minimum, compared to 40% with the global minimum solution. These differences are consistent with the numerical problems documented in the empirical literature by Metaxoglou and Knittel (2014).

2 Efficient estimation and the choice of instruments

How should the instruments be formed from the data? In the case of the linear parameters $\boldsymbol{\beta}$ the choice is natural - the vector $\boldsymbol{x}_{jt}$ of own product characteristics are the optimal instrument for $\boldsymbol{\beta}$. However, for the case of the non-linear parameters $\boldsymbol{\lambda}$ the choice is far less obvious, and can have important consequence of the performance of the estimator as we saw above.

Ai and Chen (2003) showed that it is possible to construct (asymptotically) optimal instruments directly from the data without prior knowledge of the model parameters.[14] More specifically, an efficient instrument, $A_j(\boldsymbol{x}_t)$, is defined as a vector of L dimension basis-functions such that as $L \to \infty$ and $L/n \to 0$, it is possible to approximate arbitrarily well the conditional moment restriction via a nonparametric sieve regression:

$$\begin{aligned} 0 &= E\left[\rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) | \boldsymbol{x}_t\right] \\ &= E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\lambda}\right) | \boldsymbol{x}_t\right] - \boldsymbol{x}_{jt}\boldsymbol{\beta} \\ &\approx A_j(\boldsymbol{x}_t)\hat{\boldsymbol{\pi}} - \boldsymbol{x}_{jt}\boldsymbol{\beta} \end{aligned} \tag{15}$$

where $\hat{\boldsymbol{\pi}}$ is the least-square coefficient vector obtained by regressing the inverse-demand on the vector of basis function $A_j(\boldsymbol{x}_t)$. Note that $A_j(\boldsymbol{x}_t)$ includes $\boldsymbol{x}_{jt}$.

[13] Formally, two parameter vectors are defined to be different if the smallest absolute difference is largest than 0.001. Parameter estimates obtained after failed convergence attempts are excluded from this exercise.

[14] Newey (1993) defines an instrument to be *efficient* if it achieves the semiparametric efficiency bound defined by Chamberlain (1987). The "optimal instruments" defined Chamberlain (1987) cannot be constructed directly from the data when the model in non-linear. See Newey (1990) for a two-step non-parametric procedure to estimate the optimal instruments.

Equation (15) can be thought of as an ideal "first-stage" regression and clarifies the role of the instruments in identifying the random-coefficient parameters $\boldsymbol{\lambda}$: A relevant instrument vector is set of variables characterizing the menu of characteristics available to consumers that can predict the inverse demand function of each product.

However, in our context, summarizing the information contained in $(j, \boldsymbol{x}_t)$ is a daunting task. Recall that the number of exogenous variables $\boldsymbol{x}_t$ is equal to $K \times J_t$, and the endogenous variables in the structural equation is equal to J_t. In many applications the number of products is at least as large as the number of markets/periods. This creates a curse of dimensionality problem limiting our ability to use this approach to describe the form of the optimal instruments; at least without making further restrictions on the shape of the reduced-form function that needs to be approximated.

Formally, a curse of dimensionality exists because the reduced-form of the model is a product-specific function of the entire menu of product characteristics available in the market. As the number of products in each market increases, both the number of arguments and the number of functions to approximate increase.[15] Therefore, unless the number of products is assumed to be constant and small relative to the number of markets, the number of terms necessary to approximate the function grows exponentially.[16] Intuitively, changes in market structure, from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t'}$, affect each products differentially due to market segmentation, which implies that the expectation of the inverse-demand needs to be approximated separately for each product.

To solve this curse of dimensionality, we use theoretical properties of the linear random-coefficient model to impose more structure on the reduced-form of the model. In particular, we show that the symmetry of the demand system implies that the reduced-form of the model can be written as symmetric function of the distribution of characteristic differences; a property which breaks the curse of dimensionality. We then show how this property can be used to guide the choice of the basis functions, and therefore the choice of instruments.

2.1 Symmetry of the structural and reduced-form functions

Let us define $\boldsymbol{d}_{jt,k} = \boldsymbol{x}_{jt} - \boldsymbol{x}_{kt}$ to be the vector of characteristic differences between product j and product k in market t, and let $\boldsymbol{d}_{jt} = (\boldsymbol{d}_{jt,0}, \ldots, \boldsymbol{d}_{jt,j-1}, \boldsymbol{d}_{jt,j+1}, \ldots, \boldsymbol{d}_{jt,J})$ be the matrix of differences relative to product j. Similarly, $\boldsymbol{d}_{jt}^{(2)}$ is a matrix of non-linear characteristic differences. Let us define an ordered pair $\boldsymbol{\omega}_{jt,k} = \left(s_{kt}, \boldsymbol{d}_{jt,k}^{(2)}\right)$ associated with each product $k = 0, \ldots, n_t$ in the market (including the outside good) for a given inside product $j > 0$, and let

[15] For example in the case of the original automobile data, the number of models is roughly 100 per year with 5 product characteristics. This makes $\boldsymbol{x}_t$ a 500 dimensional object. Estimating a non-parametric function of 500 variables would require an inordinate number of markets - in the BLP context there are only 20 markets (corresponding to 20 different years) and thus not even as many observations as variables.

[16] The presence of a curse of dimensionality violates Assumption 3.2 in Ai and Chen (2003), since the number of arguments in the approximating functions grow at the same rate as the sample size. However it does not affect the identification result in Berry and Haile (2014), since they implicitly assume an environment with finitely many products.

$\boldsymbol{\omega}_{jt} = (\boldsymbol{\omega}_{jt,0}, \ldots, \boldsymbol{\omega}_{jt,j-1}, \boldsymbol{\omega}_{jt,j+1}, \ldots, \boldsymbol{\omega}_{jt,J})$. We now have the following results which are proven in the Appendix A.

Proposition 1. *Under the linear in characteristics random utility model the inverse-demand*

$$\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}\right) = f\left(\boldsymbol{\omega}_{jt}; \boldsymbol{\lambda}\right) + C_t, \quad j = 1, \ldots, n_t \tag{16}$$

where C_t is a market-specific constant and f is a ***symmetric*** *function of $\boldsymbol{\omega}_{jt}$.*

The proof can be sketched as follows. We first recognize that the identity of products or the level of product attributes is irrelevant to predict consumers' discrete choice. Therefore, we can abstract from the identity of products by expressing the same demand function in terms of characteristics differences relative to product j. Furthermore, rather than normalizing the quality index of the outside good to zero, we rescale the quality index to be between zero and one: $\tau_{jt} = \exp(\delta_{jt}) / \left(1 + \sum_{j't} \exp(\delta_{j't}\right)$ for all $j = 0, \ldots, J_t$. This new normalization has the advantage of treating the outside option symmetrically with respect to the other options, and explains the presence of a market-specific intercept in equation (16). These two normalizations imply that the demand function for product j is a fully exchangeable function of the structure of the market *relative* to product j: $m_{jt} = \left\{(d^{(2)}_{jt,0}, \tau_{0t}), \ldots, (d^{(2)}_{jt,j-1}, \tau_{j-1,t}), (d^{(2)}_{jt,j+1}, \tau_{j+1,t}), \ldots, (d^{(2)}_{jt,n_t}, \tau_{n_t,t})\right\}$. The inverse mapping associated with this demand representation maintains the same symmetry and anonymity properties.

There are two key implications of Proposition 1. The first is that the inverse-demand function $\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\Sigma}\right)$ can be expressed in a fashion where it is no longer product j specific, once we condition on a vector of state variables $\boldsymbol{\omega}_{jt}$ of the products competing with j in a market.[17] The second key implication is that product invariant f is a *symmetric* function of the states of the competing products. Both of these features allow us to re-write the reduced-form equation as a symmetric function of market-structure.

To obtain this result, we need to make one more assumption.

Assumption 2. *The joint distribution of the unobserved quality of products is exchangeable in the identity of products:*

$$\Pr(\xi_{j,t} < c | \xi_{1,t}, \ldots, \xi_{j-1,t}, \xi_{j,t}, \ldots, \xi_{n_t,t}) = \Pr(\xi_{j,t} < c | \xi_{\rho(-j),t})$$

for any ordering function $\rho()$.

In economics terms, this assumption implies that the identity of rival products is not important to predict the distribution of unobservable attributes. This does not necessarily rule out the possibility that brands, for instance, are relevant for consumers' decisions. As long as brand or product

[17] Observe that the state $, \boldsymbol{\omega}_{jt,k}$ of a rival $k \neq j$ does not contain its own product characteristic $, \boldsymbol{x}_{kt}$ but rather the difference, $\boldsymbol{x}_{kt} - \boldsymbol{x}_{jt}$, relative to j.

fixed-effects enter the model linearly (shift the mean attribute), they can be concentrated out of the residual quality. This assumption is not novel in the literature. It is implicit in much of the prior empirical work, and is discussed explicitly in Berry et al. (1995) (section 5.1).

The following proposition constitutes our main theoretical result, and state that the reduced-form of the model can be written as symmetric functions of the vector of characteristic differences.

Proposition 2. *If the distribution of* $\{\xi_{1t}, \ldots, \xi_{n_t,t}\}$ *is exchangeable, the conditional expectation of the inverse-demand is a symmetric function of the matrix of characteristic differences:*

$$E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\lambda}^0\right)\Big|\boldsymbol{x}_t\right] = g\left(\boldsymbol{d}_{jt}\right) + c_t$$

where c_t *is a market specific constant.*

The proof can be sketched as follows. First, Proposition 1 implies that we can write the inverse-demand as a symmetric function by re-defining the state of the industry relative to product j. Recall that the expectation operator in equation 10 is over the market-shares vector (i.e. demand function). Since the demand for each product is symmetric, the density of shares can be re-written as a function of the entire vector of characteristics differences and the joint density of unobservable quality ξ_{jt}. This involves re-ordering the vector of characteristic differences to predict the marginal distribution of each product's market share, and does not require knowing the identity of each individual products (under Assumption 2). This establishes that the expectation of the inverse-demand is a symmetric function of the matrix $\boldsymbol{d}_{jt}$, because the joint distribution of market shares and the integrand itself are symmetric functions of characteristic differences.

Finally, a direct corollary of proposition 2 is that the optimal instruments (Chamberlain (1987), Newey (1990)) exhibit the same symmetry properties.

Corollary 1. *If the distribution of* $\{\xi_{1t}, \ldots, \xi_{n_t,t}\}$ *is exchangeable, the conditional expectation of the derivative of the residual function is a symmetric function of the matrix of characteristic differences:*

$$E\left[\left.\frac{\partial \rho_{jt}\left(\boldsymbol{s}_t, \boldsymbol{x}_t|\boldsymbol{\theta}\right)}{\partial \lambda_k}\right|\boldsymbol{x}_t\right] = g_k\left(\boldsymbol{d}_{jt}\right) + c_{t,k}, \quad \forall k = 1, \ldots, dim(\boldsymbol{\lambda})$$

where $c_{t,k}$ *is a market/parameter specific constant.*

To understand the usefulness Proposition 2 (and Corollary 1), consider a special case of the model with a single attribute, x_{jt}. In this case the state space is given by a $J_t \times 1$ vector with element k given by: $d_{jt,k} = x_{kt} - x_{jt}$. The first order polynomial approximation of the reduced-form

can written as follows:

$$\begin{aligned} g(\boldsymbol{d}_{jt}) &\approx \sum_{j' \neq j} \gamma_{j'} d_{jt,j'} \\ &= \gamma_1 \cdot \left(\sum_{j' \neq j} d_{jt,j'} \right) \end{aligned} \tag{17}$$

The second line follows directly from the symmetry of the reduced-form function. Since we can re-order the products without changing the inverse-demand (i.e. $g(\boldsymbol{d}_{jt,-j}) = g(\boldsymbol{d}_{jt,\rho(-j)})$), the coefficients of the polynomial function must be equal across products. The second order polynomial approximation takes a similar form:

$$\begin{aligned} g(\boldsymbol{d}_{jt}) &\approx \sum_{j' \neq j} \sum_{k \neq j} \gamma_{j',k} d_{jt,k} d_{jt,j'} \\ &= \gamma_1 \cdot \left(\sum_{j' \neq j} d_{jt,j'} \right) + \gamma_2 \cdot \left(\sum_{j' \neq j} \left(d_{jt,j'} \right)^2 \right) + \gamma_3 \cdot \left(\sum_{j' \neq j} d_{jt,j'} \right)^2 \end{aligned} \tag{18}$$

The main implication of Proposition 2 is therefore that one can approximate the reduced-form of the model using a small number of basis functions for which the number does not grow with the number of products available in the market, thereby breaking the curse of dimensionality.

In summary, Proposition 1 solves the curse of dimensionality in two ways. First, by expressing the state of the industry in differences (rather than in levels), it is not longer necessary to condition on the identify of products to express the inverse-demand function. This allows us too "pool" observations within and across markets since the same inverse-demand equation is used to explain the data on all products (j, t). Second, under Assumption 2, the expectation of the inverse demand is an exchangeable function of the vector of characteristics difference. This implies that the inverse-demand is function of the magnitude of characteristics, not the identity of competing products. As the previous example illustrates, this leads to a substantial reduction in the number of basis functions necessary to approximate the reduced-form.

Note finally that Proposition 1 is related to the partial-exchangeability result obtained in Pakes (1994). In particular, Pakes argues that in markets with differentiated products, a firm's demand and profit functions are partially exchangeable in its own and rivals' vector of characteristics: $\sigma_j(x_{jt}, x_{-j,t}) = \sigma_j(x_{jt}, x_{\rho(-j),t})$ for any ordering $\rho()$. While this result certainly alleviates the curse of dimensionality discussed above, it is silent in terms how x_{jt} and $x_{-j,t}$ should be interacted when constructing the basis function. In contrast, by expressing the market structure as a matrix of differences, we obtain a fully exchangeable function. This property is commonly used to alleviate the curse of dimensionality associated with the computation of Markov Perfect Equilibrium (see Doraszelski and Pakes (2007) for a survey of this literature). Farias et al. (2012) discusses various

moment-based approximation functions that exploit this property.

Implication 1: Non-Parametric Estimation

Implication 2: Differentiation IVs

Going back to the instrument selection discussion, recall that an efficient choice of instruments is a set of basis functions (e.g. polynomial of order L) that can approximate the reduced-form as L gets large. This is analogous to the suggestion of Berry et al. (1995) to use basis functions directly as instruments, rather than computing the conditional expectation of the Jacobian of the residual function as in Newey (1990). Note that Corollary 1 makes it clear that two approaches are closely related. Since the optimal instruments are symmetric functions of the same state variables, the same basis functions used as instruments in a first-stage can be used in a second-stage to approximate the optimal instruments in order improve the efficiency of the estimator.[18]

To see how Proposition 2 can be used to guide the selection of the instruments, consider the single dimension example above with a second-order polynomial approximation basis. The symmetry property restricts the number of instruments to at most three: the sum of characteristics differences, the sum of square of characteristic differences, and the square of the sum. Of course, depending of the nature of the data, it is not clear that all three should be used to construct moment restrictions, and so there is still a role for empirically choosing the optimal number of moments.

More generally, an efficient set of instruments for $\mathbf{\Sigma}$ correspond to a finite number of moments that characterize the empirical distribution of characteristic differences relative to product j in market t. Since these functions measure the degree of differentiation we label those instruments *Differentiation IVs*. To fix ideas, we consider two examples that we will use in the numerical examples in the Monte-Carlo simulations Section below. We next discuss a series of extensions and special cases in Section 2.2.

Two examples

A direct interpretation of Proposition 2 is that we can construct instrumental variables using the leading terms of the basis function of a second-order symmetric polynomial (focussing only on the binary interaction terms):

$$A_j(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} d^1_{jt,j'} \times d^l_{jt,j'}, \ldots, \sum_{j' \neq j} d^K_{jt,j'} \times d^l_{jt,j'} \right\}_{l=1,\ldots,K} \tag{19}$$

[18] A related a approach, proposed by Berry et al. (1999), is to evaluate the Jacobian of the model residual at the unconditional mean of residual (i.e. $\xi_{jt} = 0$), using preliminary estimates of the parameters (see also Goeree (2008)). Reynaert and Verboven (2013) show that this heuristic method tends to work well in practice, by creating the "right" kind of cross-sectional variation in the instruments. Our approach is complimentary to this methodology, as we discuss in more details in Section 3.5 below.

where $d^k_{jt,j'} = x_{j't,k} - x_{jt,k}$ measures the difference between product j and j' along dimension k. If the number of characteristics $K = 3$, this formulation includes 9 instrumental variables. These functions have strong economic interpretations. The sum of square of characteristic differences can be interpreted as continuous measures of distance between products, while the interaction terms capture the covariance between the various dimensions of differentiation. The Euclidian distance instrument used in Figure 1b is the square-root of the former set of instrumental variables.

Alternatively, one can exploit the symmetry property by considering only the characteristics of "close" rivals when summarizing the market structure facing each product. In most models of product differentiation (e.g. quality-ladder, hotellling, nested-logit etc), the demand for each product is most heavily influenced by a small number of alternatives with similar characteristics. For instance in a "mixed-logit quality-ladder" model, as the variance of the logit shock goes to zero, the inverse demand of product j is only function of the characteristics of products located to the right and left in the quality ranking. This feature suggests the following instrument vector:

$$A_j(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} 1\left(|d^1_{jt,j'}| < \kappa^1\right) \boldsymbol{d}_{jt,j'}, \ldots, \sum_{j' \neq j} 1\left(|d^K_{jt,j'}| < \kappa^K\right) \boldsymbol{d}_{jt,j'} \right\}, \tag{20}$$

where κ^k is a proximity threshold (e.g. standard-deviation of $x_{jt,k}$ across all markets), and $\boldsymbol{d}_{jt,j'}$ is a $K \times 1$ vector of characteristic differences between product j and j'. When characteristics are discrete, the indicator variables can be replaced by $1(d^k_{jt,j'} = 0)$.

The two formulations of the Differentiation IVs in equations (19) and (28) can include a large number of terms depending on the number of characteristics. In general, it is advisable to select a subset of those variables, based on the amount of variation across products and/or markets. For instance, it is common for some product characteristics to exhibit very little variation across markets. In Nevo (2001), the non-linear characteristics vary only at the product level (i.e. $\boldsymbol{x}^{(2)}_{jt} = \boldsymbol{x}^{(2)}_{j}$ $j = 1, \ldots, 25$), while prices vary both at the product and the market level. Assuming for simplicity that prices are exogenous (or that the researchers have valid cost shifter), one construct instruments that will be relevant to identify $\boldsymbol{\Sigma}$:

$$A_j(\boldsymbol{x}_t) = \left\{ \underbrace{\sum_{j' \neq j} 1\left(|d^1_{j,j'}| < \kappa^1\right) d^p_{jt,j'}, \ldots, \sum_{j' \neq j} 1\left(|d^K_{j,j'}| < \kappa^K\right) d^p_{jt,j'}}_{\text{Price shocks} \times \text{ Characteristic differentiation}}, \underbrace{\sum_{j' \neq j} \left(d^p_{jt,j'}\right)^2}_{\text{Price differentiation}} \right\}. \tag{21}$$

According to this formulation, the magnitude of the heterogeneity associated with market-invariant characteristic k is identified from (exogenous) variation in the relative prices of products that are more or less differentiated from product j along that particular dimension.

How does this differ from the existing literature? Interestingly, the basis function for the first-

order polynomial formulation corresponds to the suggestion in Berry et al. (1995) of using the sum of product characteristics as instruments. However, the first-order basis is collinear with market fixed-effects and the products' own characteristics. It therefore contains relatively little information to predict the inverse demand function. As we illustrated in the simulation example above, this leads to a weak identification problem of the random-coefficient parameters. In contrast, the higher-order functions discussed above vary both within and across markets, since they summarize the position of each product relative to others available in the market. Therefore, the proposed instruments can be thought of as measures of product differentiation along the exogenous dimensions of the model.

Of course the logic of differentiation IVs has been used in some applications. However, the relevance of exogenous measures of differentiation is most often justified by their ability of predict prices, rather than to identify the non-linear parameters. There exists two important exceptions: the nested-logit model (e.g. Berry (1994), Bresnahan et al. (1997)), and models of spatial differentiation (e.g. Pinkse et al. (2002), Davis (2006), Thomadsen (2007), and Houde (2012)). In both literatures, the standard instruments correspond to different versions of the proximity measures described in equation 28. From this perspective, the main contribution of this section is to formally show that the intuition developed in these two literatures remains valid in the more general random-coefficient model.

2.2 Extensions: Demographics and Price Endogeneity

Demographic variation A restrictive assumption imbedded in the derivation of the demand function in equation (4) is that the distribution of consumer preferences is common across markets. When the density function $\phi_t(\cdot|\boldsymbol{\lambda})$ is indexed by t, for instance due to variation in demographic characteristics, the reduced-form function becomes a market-specific function of the distribution of characteristic differences. There are two ways of accounting for this. First, one could specify separate moment conditions for each market. When this is not feasible, demographic characteristics can be added to the instrument vector to account for differences across markets. While previous papers have used this type of instruments, they are typically motivated as "markup" shifters (see for instance Gentzkow and Shapiro (2010) and Fan (2013)). We show that under fairly general conditions, it is feasible to transform the model so that *Differentiation IVs* analogous to the one defined above can be used to identify the non-linear preference parameters in the presence of demographic variation across markets.

To see this, consider the following single dimension example:

$$u_{ijt} = \delta_{jt} + b_{it} x_{jt}^{(2)} + \varepsilon_{ijt} \tag{22}$$

where $b_{it} = \pi_1 y_{it} + \nu_i$ is the idiosyncratic component of the marginal utility of characteristic $x_{jt}^{(2)}$ (see Nevo (2001)). This random coefficient is composed of a demographic component y_{it} that

is distributed according to (known) CDF $D_t(y)$, and a residual component ν_i that is normally distributed with mean zero and variance σ_1^2. The vector of non-linear parameters in this example contains two elements: $\boldsymbol{\lambda} = \{\pi_1, \sigma_1\}$.

Assume that the distribution of demographic characteristics can be approximated arbitrarily well using the following affine transformation of a random variable e_i:

$$y_{it} = \bar{y}_t + \text{sd}_t e_i \text{ such that } \Pr(e_i < x) = \tilde{D}(x).$$

where $\{\bar{y}_t, \text{sd}_t\}_{t=1,\ldots,T}$ and $\tilde{D}(x)$ are known transformation of the observed distribution $D_t(y)$. Of course this transformation is exact if y_{it} is normally distributed. It simply corresponds to the standardized distribution of demographic characteristics across markets.

If this assumption holds, we can write the demand function as follows:

$$\begin{aligned}
\sigma_{jt}(\boldsymbol{\delta}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\Sigma}) &= \int\int \frac{\exp\left(\delta_{jt} + \pi_1 y_{it} x_{jt}^{(2)} + \sigma_1 \eta_i x_{jt}^{(2)}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\delta_{j't} + \pi_1 y_{it} x_{j't}^{(2)} + \sigma_1 \eta_i x_{j't}^{(2)}\right)} \phi(d\eta_i) dD_t(y_{it}; \text{m}_t, \text{sd}_t) \\
&= \int \frac{\exp\left(\delta_{jt} + \sum_{k=1}^{K_2} v_{ik} \tilde{x}_{jt,k}^{(2)}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\delta_{j't} + \sum_{k=1}^{K_2} v_{ik} \tilde{x}_{jt,k}^{(2)}\right)} \psi(\boldsymbol{v}_i) d\boldsymbol{v}_i \\
&= \sigma_j(\boldsymbol{\delta}_t, \tilde{\boldsymbol{x}}_t^{(2)}; \boldsymbol{\Sigma}). \qquad (23)
\end{aligned}$$

where $\tilde{\boldsymbol{x}}_{jt}^{(2)} = \left\{\bar{y}_t x_{jt}^{(2)}, \text{sd}_t x_{jt}^{(2)}, x_{jt}^{(2)}\right\}$ is an expanded vector of non-linear characteristics, $\boldsymbol{v}_i = \{1, e_i, \eta_i\}$, and is the joint density of $\boldsymbol{v}_i$ defined from $\phi(\cdot)$ and $\tilde{D}(\cdot)$. Note that the change of variables in equation (23) allows us to eliminate the t subscript from the demand function, and expand the state space by adding two new interactions: (i) the mean of y_{it} times $x_{jt}^{(2)}$, and (ii) the standard-deviation of y_{it} times $x_{jt}^{(2)}$. Under this new parametrization of the model, we can use directly Proposition 2 to write the reduced-form of the model as follows:

$$E\left[\sigma_{jt}^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\Sigma}) | \boldsymbol{x}_t\right] = g\left(\boldsymbol{d}_{jt}, \mathbf{m}_t \otimes \boldsymbol{d}_{jt}^{(2)}\right) + c_t \qquad (24)$$

where $\mathbf{m}_t = \{\bar{y}_t, \text{sd}_t\}$ is a 1×2 vector containing the mean and standard-deviation of y_{it} in market t, and $\mathbf{m}_t \otimes \boldsymbol{d}_{jt}^{(2)}$ is the interaction of those moments with the non-linear characteristic differences. Therefore, when demographic variation is incorporated in the model in this fashion, the reduced-form of the model is a symmetric function of the same arguments as before, plus moments of the distribution of demographic characteristics interacted with the distribution of characteristics differences.

We can use this insight to construct instrument variables that identify separately the two sources of heterogeneity in the taste for characteristic $x_{jt}^{(2)}$. For instance, the first Differentiation IV example

above can be expanded by adding an interaction with demographic moments:

$$A_j(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} d^1_{jt,j'} \times d^l_{jt,j'}, \ldots, \sum_{j' \neq j} d^K_{jt,j'} \times d^l_{jt,j'}, \sum_{j' \neq j} \mathbf{m}_t \otimes \left(d^l_{jt,j'} \times d^{(2)}_{jt,j'} \right) \right\}_{l=1,\ldots,K} . \tag{25}$$

Focussing on the quadratic term (i.e. $d^l_{jt,j'} = d^{(2)}_{jt,j'}$), the added instruments capture how product differentiation asymmetrically impacts the inverse-demand of product j based on the mean and variance of consumer characteristics across markets.

The argument above holds exactly if the distribution of demographic characteristics can be standardized, and the logic can be extended to multiple dimensions. More general distributions, we need to rely on a heuristic approximation, but the logic should hold more generally:

$$E\left[\sigma_{jt}^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\Sigma}) | \boldsymbol{x}_t\right] \approx g\left(\boldsymbol{d}_{jt}, \mathbf{m}_t \otimes \boldsymbol{d}^{(2)}_{jt}\right) + c_t \tag{26}$$

where $\mathbf{m}_t$ is now a vector of moments characterizing the joint distribution of demographic characteristics in market t. The key insight of this transformation is that demographic characteristics should enter the instrument vector as interaction terms with other measures differentiation, rather than as stand-alone variables.[19]

Endogeneous prices Incorporating endogenous prices into the model does not fundamentally change the identification problem of $\boldsymbol{\Sigma}$, but adds an additional simultaneity problem: in equilibrium prices are correlated with the unobserved quality of products (Berry et al. 1995).

To see how this a new source of simultaneity changes the problem, consider the following slight change of notation to the inverse-demand:

$$\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t | \boldsymbol{\lambda}^0\right) = f\left(\boldsymbol{\omega}_{jt}\right) + C_t.$$

Element k of the state vector $\boldsymbol{\omega}_{jt}$ now includes: $\{s_{kt}, \boldsymbol{d}^{(2)}_{jt,k}, \boldsymbol{d}^{p}_{jt,k}\}$, where $d^p_{jt,k}$ is the price differences between product j and k. As before, $f(\cdot)$ is a symmetric function of industry state vector $\boldsymbol{\omega}_{jt}$. Although the conditional expectation of equation 27 is also a symmetric function, the conditional mean restriction in equation 10 is no longer satisfied:

$$E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}^0\right) | \boldsymbol{x}_t, \boldsymbol{p}_t\right] - \boldsymbol{x}_{jt}\boldsymbol{\beta} = g(\boldsymbol{d}_{jt}, \boldsymbol{d}^p_{jt}) + c_t - \boldsymbol{x}_{jt}\boldsymbol{\beta} \neq 0.$$

Two sources of variation have been exploited in the literature to construct valid price instruments: (i) ownership structure (e.g. Berry et al. (1995)), and (ii) cost-shifters (e.g. Nevo (2001)).

[19] See Romeo (2010) for a similar argument and simulation results showing the importance of accounting for interactions between product characteristics and the mean of demographic attributes in the instrument vector.

Let $\boldsymbol{w}_{jt}$ denotes a vector of relevant and valid price instruments such that:

$$E[\xi_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t] = 0. \tag{27}$$

The challenge in using this restriction to construct instruments, is that the reduced-form of the model cannot be written as a symmetric function of $\{\boldsymbol{d}_{jt}, \boldsymbol{d}^w_{jt}\}$. To see this, recall that the symmetry of the reduced-form arises from the symmetry of the demand function itself (since the conditional expectation is over $\boldsymbol{s}_t$). With endogenous prices, the conditional expectation of the inverse demand depend also on the distribution of prices given $(\boldsymbol{x}_t, \boldsymbol{w}_t)$ (i.e. firms' conduct). This distribution is unlikely to be a symmetric function of characteristic differences, since the identity/ownership of products is expected to play an important role. It is therefore impossible to obtain an exact characterization of the reduced-form without knowing the exact supply relation governing prices. Importantly, this does not imply that it is not feasible to construct valid/relevant instruments in this case. It just means that we cannot provide a exact characterization of the reduced-form of the model that solves the curse of dimensionality problem for all supply models.

To get around this problem, we rely on a heuristic approximation of the reduced-formed first proposed by Berry et al. (1999), and recently reexamined by Reynaert and Verboven (2013). The argument proceed in two steps.

First, let $\hat{p}_{jt} \approx E(p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t)$ denotes an estimate of the reduced-form of the pricing equation constructed from observed characteristics. This *exogenous* price measure can be constructed using linear regressions by exploiting random variation from cost or ownsfership shocks (as in Reynaert and Verboven (2013)), or by solving an equilibrium pricing game after setting $\xi_{jt} = 0$ (as in Berry et al. (1999)). The choice of the approach is application/data specific. Since $\hat{p}_{jt}$ is constructed from $(\boldsymbol{x}_{jt}, \boldsymbol{w}_{jt})$, the following conditional moment restriction is satisfied:

$$E\left[\xi_{jt}|\boldsymbol{x}_t, \hat{\boldsymbol{p}}_t\right] = 0.$$

Second, following insights from Berry et al. (1999), we use the following heuristic approximation of the reduced-form:

$$\begin{aligned} E_{p,s}\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}^0\right)|\boldsymbol{x}_t, \boldsymbol{w}_t\right] &\approx E_s\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \hat{\boldsymbol{p}}_t; \boldsymbol{\lambda}^0\right)|\boldsymbol{x}_t, \hat{\boldsymbol{p}}_t\right] \\ &= g\left(\boldsymbol{d}_{jt}, \boldsymbol{d}^{\hat{p}}_{jt}\right) + c_t \end{aligned}$$

The heuristic approximation distributes the expectation operator over prices inside the non-linear function $\sigma_j^{-1}(\cdot)$. The second equality follows from the fact that after replacing p_{jt} with $\hat{p}_{jt}$ in the inverse-demand, we obtain a reduced-form representation that is a symmetric function of $\boldsymbol{d}^{\hat{p}}_{jt}$.

Berry et al. (1999) use the same heuristic to approximate the optimal instruments. The difference between our two approaches, is that we use this approximation to approximate the

conditional expectation over prices, while Berry et al. (1999) use it to approximate the conditional expectation over prices and shares. Reynaert and Verboven (2013) show that this heuristic approach works well in practice for models with a small to medium number of random coefficients. See Section 3.5 below for a comparative analysis of the two approaches.

As before, we can then construct instrument functions that exploit the symmetry of $g(\boldsymbol{d}_{jt}, \boldsymbol{d}_{jt}^{\hat{p}})$. For instance, the "local" differentiation IVs example above becomes:

$$A_j(\boldsymbol{x}_t, \boldsymbol{w}_t) = \left\{ \hat{p}_{jt}, \sum_{j' \neq j} 1\left(|d_{jt,j'}^1| < \kappa^1\right) \boldsymbol{d}_{jt,j'}, \ldots, \sum_{j' \neq j} 1\left(|d_{jt,j'}^K| < \kappa^K\right) \boldsymbol{d}_{jt,j'}, 1\left(|d_{jt,j'}^{\hat{p}}| < \kappa^{\hat{p}}\right) \boldsymbol{d}_{jt,j'} \right\}, \quad (28)$$

where the vector characteristic differences $\boldsymbol{d}_{jt,j'}$ is expanded to included (exogenous) price differences $d_{jt,j'}^{\hat{p}}$. Note that $\hat{p}_{jt}$ is included in $A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)$ to instrument for p_{jt}.

3 Monte-Carlo Simulations

In this section, we analyze the finite sample properties of the Differentiation IVs described in the previous section. We consider four models: (i) exogenous characteristics with independent random-coefficients, (ii) exogenous characteristics with correlated random-coefficients, and (iii) endogenous prices, and (iv) endogenous product locations. We illustrate how to construct relevant instruments for each of these cases. In the last subsection, we compare our IV approach to the optimal IV approximation proposed by Berry et al. (1999) and Reynaert and Verboven (2013).

3.1 Independent random-coefficients

We start by revisiting the numerical example discussed in Section 1.3:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + \sum_{k=1}^{K_2} (\beta_{2,k} + \lambda_k \eta_{ik}) \cdot x_{jt,k}^{(2)} + \xi_{jt} + \epsilon_{ijt}, \quad j = 1, \ldots, 15 \text{ and } t = 1, \ldots, 100$$

where $\eta_{ik} \sim N(0,1)$ and $\epsilon_{ijt} \sim \text{T1EV}(0,1)$. See footnote 7 for a description of the data-generating process. We simulate models with number of random of random coefficients ranging from 1 to 7 (i.e. K_2). For each specification, we compare the performance two differentiation IVs:

$$\text{Quadratic Diff IV:} \qquad \boldsymbol{z}_{jt} = \left\{ \boldsymbol{x}_{jt}, \sum_{j'} \left(d_{jt,j'}^1\right)^2, \ldots, \sum_{j'} \left(d_{jt,j'}^K\right)^2 \right\}$$

$$\text{Local Diff IV:} \qquad \boldsymbol{z}_{jt} = \left\{ \boldsymbol{x}_{jt}, \sum_{j'} 1\left(|d_{jt,j'}^1| < \text{sd}_1\right), \ldots, \sum_{j'} 1\left(|d_{jt,j'}^1| < \text{sd}_K\right) \right\}$$

Table 2: Monte-Carlo simulation results for exogenous characteristics model with strong instruments

(a) Differentiation IV: Quadratic

	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\log \lambda_1$	0.000	0.030	-0.000	0.032	-0.001	0.033	-0.001	0.036
$\log \lambda_2$			-0.002	0.031	0.000	0.032	-0.002	0.035
$\log \lambda_3$					-0.000	0.031	-0.001	0.034
$\log \lambda_4$							-0.002	0.036
λ_1	0.002	0.122	0.001	0.130	-0.003	0.133	-0.003	0.142
λ_2			-0.004	0.125	0.004	0.128	-0.004	0.141
λ_3					0.001	0.125	-0.001	0.137
λ_4							-0.005	0.146
1(Local)	0.000		0.000		0.000		0.000	
Rank-test – $F(1)$	1202.104		564.033		330.399		206.417	
p-value	0.000		0.000		0.000		0.000	
IIA-test – $F(K)$	359.409		363.224		321.730		276.135	
p-value	0.000		0.000		0.000		0.000	

(b) Differentiation IV: Local

	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\log \lambda_1$	-0.000	0.032	-0.001	0.034	-0.000	0.034	-0.001	0.037
$\log \lambda_2$			-0.002	0.032	0.000	0.033	-0.001	0.037
$\log \lambda_3$					-0.001	0.033	-0.001	0.037
$\log \lambda_4$							-0.003	0.038
λ_1	0.002	0.126	0.000	0.135	0.002	0.137	-0.003	0.147
λ_2			-0.007	0.128	0.004	0.131	-0.003	0.148
λ_3					-0.002	0.130	-0.001	0.148
λ_4							-0.008	0.152
1(Local-min)	0.000		0.000		0.000		0.000	
Rank-test – $F(1)$	1050.015		523.760		322.288		204.402	
p-value	0.000		0.000		0.000		0.000	
IIA-test – $F(K)$	297.544		298.073		262.636		222.932	
p-value	0.000		0.000		0.000		0.000	

where K is the number of characteristics (excluding the intercept), and sd_k is the standard-deviation of $x_{jt,k}$. Tables 3 and 2b summarize the simulation results.

Table 3 mirrors the results obtained in Table 1 with weak instruments (i.e. sum of rival characteristics). Both differentiation IVs specifications allow us to reject the null hypothesis of under-

Table 3: Small sample and asymptotic distribution for exogenous characteristic model with strong instruments

	Diff IV: Quadratic			Diff IV: Local		
	bias	rmse	asym-se	bias	rmse	asym-se
$K_2 = 1$	0.000	0.030	0.031	-0.000	0.032	0.032
$K_2 = 2$	-0.001	0.032	0.031	-0.001	0.033	0.032
$K_2 = 3$	-0.000	0.032	0.033	-0.000	0.033	0.034
$K_2 = 4$	-0.001	0.035	0.035	-0.002	0.037	0.036
$K_2 = 5$	0.000	0.039	0.039	-0.000	0.040	0.040
$K_2 = 6$	-0.001	0.045	0.044	-0.001	0.046	0.045
$K_2 = 7$	0.002	0.048	0.050	-0.003	0.051	0.052

identification (rank-test), as well as the IIA hypothesis. In addition, the frequency of local optima is equal to zero across all specifications; meaning that the numerical optimization algorithm always converged to the estimates irrespectively of the starting values.

The precision and bias of the parameter estimates are also small across all specifications. The RMSE for $\hat{\sigma}_x$ are roughly 17 times larger with the strong instruments Table 3, compared with the weak instruments used in Table 1.

Table 2b summarizes the precision of the estimates across all specifications, including the average asymptotic standard-errors. Consistent the discussion above, we see that loss in precision from adding random-coefficients is very minor. The RMSE increases from 0.03 to 0.05 when we increase the number of random-coefficients to seven. This is encouraging since the sample size is fairly small: 15 products $\times$ 100 markets. In addition, the asymptotic standard-errors are nearly identical to the RMSE across all specifications, consistent with the idea that the two alternative Differentiation IVs are strong in this context.

Finally, note that at least in this example, the quadratic instruments appear to perform marginally better than the local differentiation IVs. This is reflected in slightly smaller RMSE and biases, and largest Rank-test and IIA-test statistics.

3.2 Correlated random-coefficients

Next, we consider a model with exogenous characteristics in which consumers have correlated random-coefficients:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + \sum_{k=1}^{K_2} (\beta_{2,k} + \nu_{ik}) \cdot x_{jt,k}^{(2)} + \xi_{jt} + \epsilon_{ijt}, \quad j = 1, \ldots, 50 \text{ and } t = 1, \ldots, 100, \tag{29}$$

where $\boldsymbol{\nu}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, and $K_2 = 4$. We use a larger sample for this example: $J_t = 50$ instead $J_t = 15$. This reflects the fact that the number of non-linear parameters is substantially larger with correlated random-coefficients: from 4 to 10. The sample size is still relatively small compared to

Table 4: Variance-Covariance matrix of random-coefficients

(a) Variance-Covariance Matrix ($\boldsymbol{\Sigma}$)

	$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$	$\Sigma_{\cdot,4}$
$\Sigma_{1,\cdot}$	4			
$\Sigma_{2,\cdot}$	-2	4		
$\Sigma_{3,\cdot}$	2	-2	4	
$\Sigma_{4,\cdot}$	2	-2	2	4

(b) Choleski Decomposition

	$C_{\cdot,1}$	$C_{\cdot,2}$	$C_{\cdot,3}$	$C_{\cdot,4}$
$C_{1,\cdot}$	1.99	0	0	0
$C_{2,\cdot}$	-1	1.73	0	0
$C_{3,\cdot}$	1	-0.58	1.63	0
$C_{4,\cdot}$	1	-0.58	0.48	1.58

what applied researchers typically used.

Table 4 reproduces the variance-covariance used in the simulations. The variance terms are the same as the ones in the previous specification. The covariance terms are chosen such that the correlation between random-coefficients are either -0.5 or 0.5.

Note that we do not estimate $\boldsymbol{\Sigma}$ directly, but the Choleski decomposition of $\boldsymbol{\Sigma} = \boldsymbol{C}'\boldsymbol{C}$. This allows us to write indirect utility of consumers as a linear function of parameters and K_2 standard-normal random-variables: $\boldsymbol{\nu}_i = \boldsymbol{C}'\boldsymbol{\eta}_i$ where $\boldsymbol{\eta}_i \sim \mathcal{N}(0, I)$. To ensure that $\boldsymbol{\Sigma}$ is positive semidefinite, we constraint the diagonal elements of $\boldsymbol{C}$ to be positive by estimating the log of $C_{k,k}$.

To construct an instrument vector, we use the quadratic form of the Differentiation IVs with additional interaction terms between each characteristics pairs:[20]

$$\boldsymbol{z}_{jt} = \left\{ x_{jt}, \sum_{j' \neq j} d^1_{jt,j'} \times d^1_{jt,j'}, \ldots, \sum_{j' \neq j} d^l_{jt,j'} \times d^l_{jt,j'}, \sum_{j' \neq j} d^1_{jt,j'} \times d^{l+1}_{jt,j'}, \ldots, \sum_{j' \neq j} d^K_{jt,j'} \times d^K_{jt,j'} \right\}. \quad (30)$$

This results in 15 excluded restrictions: (i) five quadratic differentiation measures along each dimension (one special regressors and four non-linear characteristics), and (ii) ten unique interaction pairs.

The simulation results are summarized in Table 5. The top panel reports the average estimated parameters (transformed) of the variance-covariance matrix, the middle panel reports the RMSE associated with each parameter, and the bottom panel reports the averages of the IIA-test and the Cragg-Donald rank test statistics. Both tests confirm once again that the instruments are strong, and that the IIA hypothesis is easily rejected. The average bias and RMSE are also remarkably small, despite the richness of the model. The bias and precision of the Choleski parameter estimates are reproduced in the last three columns of Table 8 below. The differentiation IVs are able to accurately identify both the magnitude and correlation in taste heterogeneity across consumers.

It is worth noting that to our knowledge this specification is substantially richer than any random-coefficient model that has previously been studied with aggregate data by researchers,

[20] Similar interactions can be constructed with the local differentiation instruments: $\sum_{j'} 1(|d^l_{jt,j'}| < \kappa_l) d^k_{jt,j'}$. The results are similar using this specification of the instruments, but we find that the quadratic form tends to be more stronger.

Table 5: Monte-Carlo simulation results for correlated random-coefficient specification with differentiation IVs

		$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$	$\Sigma_{\cdot,4}$
Estimates	$\Sigma_{1,\cdot}$	4.003			
	$\Sigma_{2,\cdot}$	-1.997	4.000		
	$\Sigma_{3,\cdot}$	1.997	-1.996	3.991	
	$\Sigma_{4,\cdot}$	2.010	-2.000	2.006	4.010
RMSE	$\Sigma_{1,\cdot}$	0.228			
	$\Sigma_{2,\cdot}$	0.132	0.232		
	$\Sigma_{3,\cdot}$	0.156	0.145	0.217	
	$\Sigma_{4,\cdot}$	0.156	0.143	0.154	0.217
IIA test (F)		157.637			
Cragg-Donald statistic (F)		474.053			
Nb endogenous variables		10			
Nb IVs		15			

both in empirical applications and Monte-Carlo simulations. Although we obtain these results in a "controlled" environment, this result confirms the idea in Berry, Levinsohn, and Pakes (1995) and Berry and Haile (2014) that it is feasible to estimate very flexible substitution patterns using aggregate data on market shares and product availability.

3.3 Endogenous prices

To study the performance of the Differentiation IVs when prices are endogenous we consider a model with a single random-coefficient on price:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + (\beta_p + \lambda_p \nu_i) \cdot p_{jt} + \xi_{jt} + \epsilon_{jt}, \quad j, = 1, \ldots, 15 \text{ and } t = 1, \ldots, 100. \tag{31}$$

where $\ln \nu_i \sim \mathcal{N}(0,1)$.[21] We use the following parameter values in the simulation: $\beta_0 = 50$, $\beta_x = 2$, $\beta_p = -0.2$ and $\lambda_p = -4$.

To generate a second simultaneity problem, we generate prices using a Bertrand-Nash pricing game with single-product competitors:

$$p_{jt}^* = c_{jt} - \sigma_j(\boldsymbol{\delta}_t, \boldsymbol{p}_t^*; \lambda_p) \left[\frac{\partial \sigma_j(\boldsymbol{\delta}_t, \boldsymbol{p}_t^*; \lambda_p)}{\partial p_{jt}^*} \right]^{-1} \tag{32}$$

$$\text{Where,} \qquad c_{jt} = \gamma_0 + x_{jt}^{(1)} \gamma_x + \omega_{jt}.$$

The marginal-cost function is assumed to be constant, and the cost-shock ω_{jt} is observed by the

[21] Unlike the previous examples, we approximate the distribution of ν_i using a fixed sample of 100 pseudo random-numbers.

econometrician. We use this variable below to construct a price instrument. The data is generated by repeatedly finding a solution to equation (32) for 1000×100 markets.[22] This gives us $1,000$ independent panels to conduct our Monte-Carlo simulation analysis.

We follow the steps described in Section 2.2 to construct the Differentiation IVs. We first construct an exogenous price index, $\hat{p}_{jt}$, using the predicted values from the linear regression of p_{jt} on the exogenous characteristic and the cost shifter ω_{jt}:

$$\hat{p}_{jt} = \hat{\pi}_0 + \hat{\pi}_1 x_{jt}^{(1)} + \hat{\pi}_2 \omega_{jt}. \tag{33}$$

We then construct the Differentiation IVs using the empirical distribution of differences in $\hat{p}_{jt}$ and $x_{jt}^{(1)}$. In particular, as before, we consider the two alternative specifications:

$$\text{Quadratic Diff IV:} \qquad \boldsymbol{z}_{jt} = \left\{ \boldsymbol{x}_{jt}, \omega_{jt}, \sum_{j'} \left(d_{jt,j'}^{(1)} \right)^2, \sum_{j'} \left(d_{jt,j'}^{\hat{p}} \right)^2 \right\}$$

$$\text{Local Diff IV:} \qquad \boldsymbol{z}_{jt} = \left\{ \boldsymbol{x}_{jt}, \omega_{jt}, \sum_{j'} 1\left(|d_{jt,j'}^{(1)}| < \text{sd}_1 \right), \sum_{j'} 1\left(|d_{jt,j'}^{\hat{p}}| < \text{sd}_{\hat{p}} \right) \right\}$$

where $d_{jt,j'}^{(1)} = x_{j't}^{(1)} - x_{jt}^{(1)}$ and $d_{jt,j'}^{\hat{p}} = \hat{p}_{j't} - \hat{p}_{jt}$. Note that ω_{jt} is added to the list of instruments since we need two independent sources of variation to identify β_p and λ_p (i.e. own cost shifters, and cost and characteristics of rival products).

The simulation results are reproduced in Figure 3 and Table 6. In addition to the two sets of instruments defined above, we also report the results using the "sum of rival characteristics" in order to illustrate effect weak instruments in this setting with two sources of simultaneity.

Table 6b confirm the results obtained with the weak instrument specification in Section 1.3. The test statistics associated the IIA hypothesis and Cragg-Donald rank-test are on average significantly below standard critical values.[23] Recall that the local identification condition by estimating the reduced-form of the model:

$$\begin{aligned} p_{jt} &= \boldsymbol{z}_{jt}\boldsymbol{\pi}_1 + e_{jt}^1 \\ \frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t; \lambda_p^0)}{\partial \lambda_p} &= \boldsymbol{z}_{jt}\boldsymbol{\pi}_2 + e_{jt}^2 \end{aligned}$$

To illustrate the dual role of the instruments in this context, we report the results of two first-stage F tests: (i) one that simply regress price and the Jacobian on the exogenous variables, and

[22] The data-generating process for the marginal cost and characteristics is given by: $\xi_{jt} \sim \mathcal{N}(0,1)$, $x_{jt}^{(1)} \sim \mathcal{N}(0,1)$, $\omega_{jt} \sim \mathcal{N}(0, 0.1)$.

[23] Since the moments are weak, we use the Stock-Yogo critical values to evaluate the Cragg-Donald statistics (Stock and Yogo (2005)). As we discussed in Section 1.3, this rank-test is valid in our non-linear context since we evaluate the Jacobian function at the true value of the parameters.

Table 6: Monte-Carlo simulation results for endogenous price specification

(a) Distribution of parameter estimates

	(1)	(2)			(3)			(4)		
	True	Diff. IV = Local			Diff. IV = Quadratic			Diff. IV = Sum		
		bias	se	rmse	bias	se	rmse	bias	se	rmse
λ_p	-4.00	0.02	0.27	0.28	0.02	0.53	0.55	1.03	158.25	2.10
β_p	-0.20	0.01	0.37	0.37	0.01	0.31	0.32	-0.67	201.29	1.38
β_0	50.00	-0.26	3.92	3.92	-0.28	7.36	7.45	-9.82	26.41	20.65
β_x	2.00	-0.02	0.46	0.45	-0.02	0.47	0.47	0.34	1.11	0.83

(b) Weak identification tests

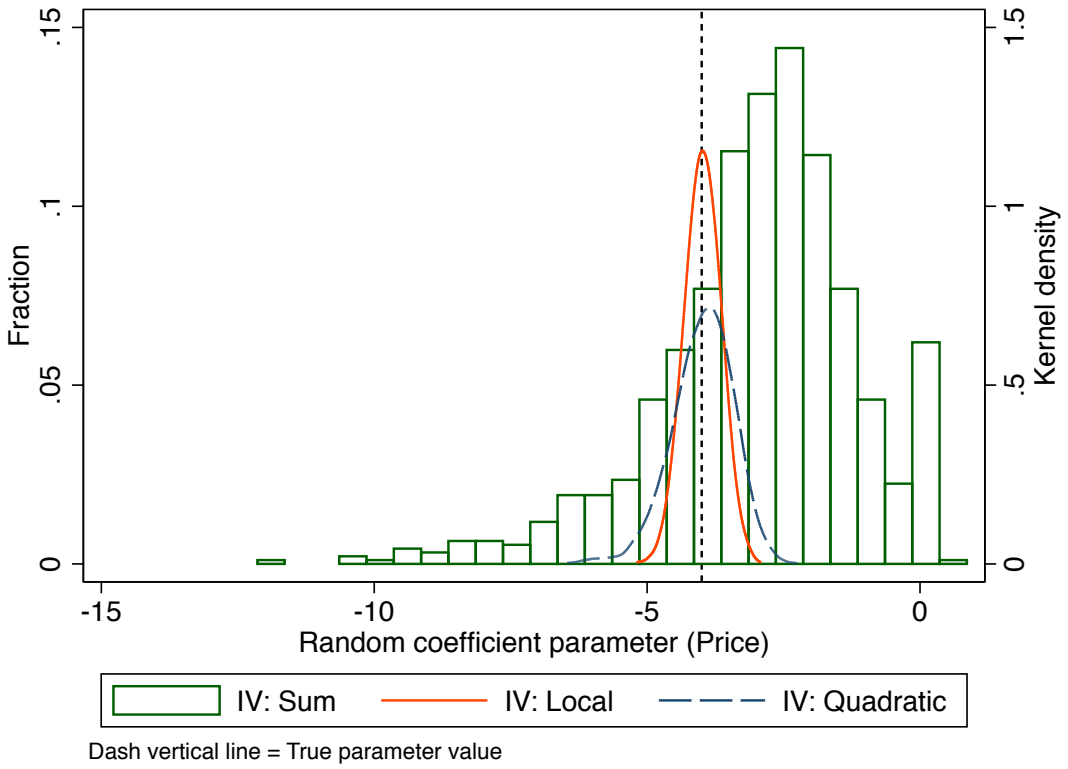
	(1)	(2)	(3)
	IV = Local	IV=Quadratic	IV = Sum
Frequency conv.	1	1	0.94
IIA-test	109.48	53.90	1.88
p-value	0	0	0.34
1st-stage F-test: Price	191.80	442.10	138.94
1st-stage F-test: Jacobian	214.60	58.40	27.85
Cond. 1st-stage F-test: Price	252.23	479.96	7.92
Cond. 1st-stage F-test: Jacobian	280.31	82.44	6.19
Cragg-Donald statistics	170.19	54.45	4.09
Stock-Yogo size CV (10%)	16.87	13.43	13.43
Nb. endogenous variables	2	2	2
Nb. IVs	4	3	3

(ii) one that first "project-out" the variation of the other endogenous variable before computing the first-stage F test. The second test was proposed by Angrist and Pischke (2009) and Sanderson and Windmeijer (2016) to adjust the standard F-tests for IV cases with multiple endogenous variables. The results highlight the importance of correctly measuring the weakness of the moment restrictions as a whole.

In our example, the standard F-tests conducted using the sum of rival characteristics incorrectly suggest that weak instruments is not a concern (i.e. 138.94 and 27.85). This is because one of the instrument is very strong (i.e. cost shifter ω_{jt}). However, once we account for the fact that we have more than one parameters to identify with a single strong instrument, the conditional first stage F-tests are in line with the results of the Cragg-Donald and the IIA tests; both F-tests are significantly below the Stock-Yogo critical values on average.

Next, looking at columns (1) and (2) in Table 6b we see that both instruments constructed using measures of differentiation eliminate the weak identification problem. This time however, the "Local Differentiation IV" tends to perform significantly better than the "Quadratic Differentiation

Figure 3: Distribution of estimated price random-coefficient parameter for alternative differentiation instruments



IV". The two measures of weakness, the IIA-test and the Cragg-Donald statistic, are on average roughly 2.5 times larger in column (1) than in column (2).

Table 6a summarizes the distribution of the estimated parameters across the three IV specifications. Looking first at the weak IV panel, we see again that using weak instruments lead to substantial loss in precision and large biases. The RMSE for λ_p is equal 2.10, and the average bias is significantly above zero (1.03). This upward bias is partial offset by a "downward" bias in β_p (i.e. -0.67), but the net effect is positive: weak instruments in this example biases the slope of the demand towards zero.

This bias is eliminated in panel (2) and (3) when we use the stronger differentiation IVs. The RMSEs are also substantially reduced. Relative to the sum of rival characteristics specification, we obtain a 7.5 improvement in precision for $\hat{\sigma}_p$ with the local differentiation IV, and a 4 times improvement with the quadratic differentiation IV.

Figure 3 illustrates this point graphically by plotting the distribution of $\hat{\sigma}_p$ for the three specifications. As with the exogenous characteristics, weak instruments lead to a non-gaussian distribution of the parameters, characterized by large outliers and a mass around zero. The two other dis-

tributions are symmetric and bell-shape, centered around the true parameter, and do not exhibit outliers. The comparison between the spread of the distributions with "Quadratic" and "Local" Differentiation IVs also illustrate the gain in precision from using a stronger set of moments.

It is unclear why the local differentiation measure performs better the quadratic measure in this example. The conditional first-stage test for price in Table 6b suggests that it is not the case that the number of local competitors is a better predictor of prices; if anything prices appear to be *more* correlated with the sum of square of characteristic differences (i.e. $F = 479.96$ versus $F = 252.23$). We conjecture that the performance difference is due to two factors. First, unlike in the first example, the distribution of the random-coefficient is not symmetric (due to the log-normality assumption), and so it is possible that the second-order polynomial is not an accurate approximate the reduced-form of the model. Second, it is important to note that the heuristic approximation associated with using $\hat{p}_{jt}$ instead of p_{jt}, introduces measurement error in the instruments. The sum of square of $d^{\hat{p}}_{jt,j'}$ tends to amply this error since the instrument is proportional to the variance of $\hat{p}_{jt}$. In contrast, measurement error in $\hat{p}_{jt}$ leads to a random mis-classification of firms between "local" and "non-local" competitors, but does not systematically bias the number of local competitors. The quadratic differentiation IV is therefore more sensitive to measurement error, which could explain why those moments are weaker on average.

3.4 Endogenous Characteristics and Natural Experiments

An often expressed criticism of the main identifying assumption in Berry et al. (1995), is that firms endogenously choose product characteristics (observed *and* unobserved) in response to changes in the structure of the market. This violates Assumption 1 either because of an endogenous selection of products, and/or because of the a contemporaneous correlation between ξ_{jt} and the attributes of own and rival products.[24] This invalidates the use of the entire distribution of characteristic differences to identify substitution patterns.

An alternative approach is to look for *natural experiments* that exogenously change the menu of product characteristics available to consumers. Such experiments can be induced directly by researchers (e.g. Conlon and Mortimer (2015)), caused by exogenous technology changes that generate the shakeout of an industry (e.g. Houde (2012)), or by government regulations that generate suboptimal product offering (e.g. zoning). To illustrate this, consider the following mixed-logit Hotelling demand model:

$$u_{ijmt} = \begin{cases} \xi_{jmt} - \lambda(\nu_i - x_{jmt})^2 + \epsilon_{ijmt} & \text{If } j > 0, \\ \epsilon_{ijmt} & \text{If } j = 0. \end{cases}$$

where $j = 1, \ldots, 15$ indexes products, $m = 1, \ldots, 100$ indexes markets, and $t = 0$ or 1 indexes the

[24] See Ciliberto et al. (2016) for a recent examination of this problem.

pre/post natural experiment periods. In this example, the non-linear characteristic of products, x_{jmt}, measures their location in the product space, and while the random-coefficient, ν_i, measures the "ideal" address of consumers. We assume that both variables are uniformly distributed between 0 and 2. The goal is to estimate the travel cost of consumers: λ.

We consider a natural experiment associated with the entry of a new product in each market at location x^* in the post-period (i.e. $t = 1$). Within each market, distance to x^* measures the strength of the "treatment". The characteristics of incumbent products are constant across periods.

We introduce a correlation between ξ_{jmt} and $\boldsymbol{x}_m$ as follows:

$$E(\xi_{jmt}) = 0 \text{ and } \text{corr}(\xi_{jmt}, \text{ED}_{jm}) = a < 0$$

where $\text{ED}_{jm} = \sqrt{\sum_{j'}(x_{jm} - x_{j'm})^2}$ is the Euclidian distance of incumbent product j. The parameter a creates a standard simultaneity problem: products facing close substitutes have higher unobserved quality. Since characteristics are constant across the two periods, this correlation can be absorbed by conditioning on product/market fixed-effects. Assumption 3 formalizes this assumption.

Assumption 3. *The change in the unobserved quality of products has mean zero conditional on the observed menu of characteristics:*

$$E\left[\Delta\xi_{jm} | \boldsymbol{x}_{mt}\right] = 0, \forall t = 0, 1$$

where $\Delta\xi_{jm} = \xi_{jm1} - \xi_{jm0}$.

To construct the instruments, we consider two distance measures similar to the Differentiation IVs discussed above:

$$\begin{aligned} w^1_{jm} &= 1(|x_{jm} - x^*| < \kappa) \\ w^2_{jm} &= (x_{jm} - x^*)^2 \end{aligned}$$

where the threshold κ is defined as the standard deviation of x_{jm} across all products/markets. Let $\boldsymbol{z}_{jm} = \{1, w^1_{jm}, w^2_{jm}\}$ denotes the instrument vector. This leads to the following moment condition:

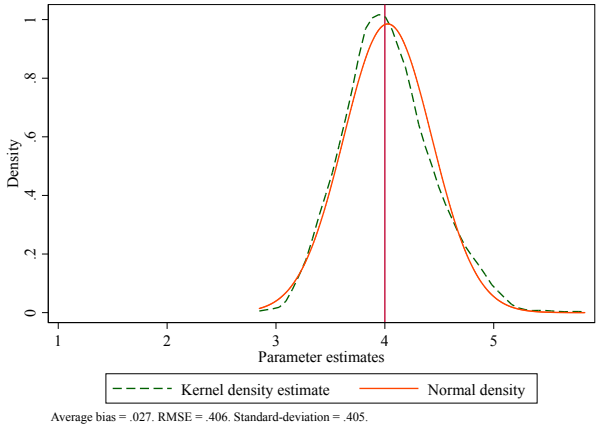
$$g_n(\lambda) = \frac{1}{n}\sum_m \sum_j \left[\rho_j(\boldsymbol{s}_{m1}, \boldsymbol{x}_{m1}; \lambda) - \rho_j(\boldsymbol{s}_{m0}, \boldsymbol{x}_{m0}; \lambda)\right] \cdot \boldsymbol{z}_{jm} = \Delta\boldsymbol{\rho}(\lambda)^T \boldsymbol{z}/n$$

where n is the number of unique market/product observations.

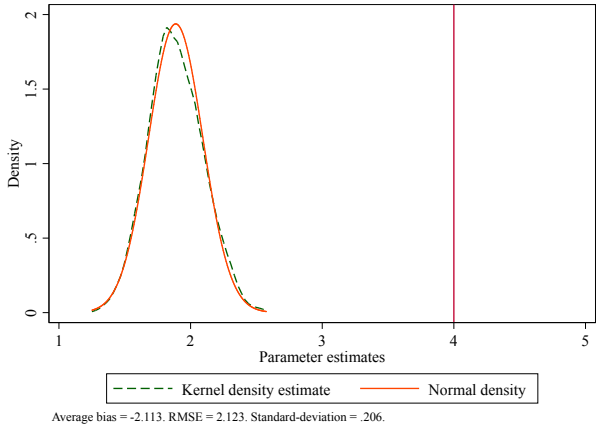
Using this specification, the structural parameters of the model are identified solely from the quasi-experimental variation. In particular, the reduced-form of the GMM problem corresponds to a difference-in-difference regression. To see this, note that the Gauss-Newton regression associated

Figure 4: Monte-Carlo simulated distribution of the travel cost parameter estimates with endogenous product locations

(a) Difference-in-Difference



(b) Differentiation IVs



Data generating process: $x_{jm} \sim U[0,2]$, $\xi_{jmt} = \bar{\xi}_{jm} + \Delta\xi_{jmt}$, where $\bar{\xi}_{jm} = -0.25\left(\text{ED}_{jm} - \overline{\text{ED}}_m\right) + \zeta_{jm}, \zeta_{jm} \sim N(0, 0.5)$ and $\Delta\xi_{jmt} \sim N(0, 0.25)$. Consumer addresses: $\nu_i \sim U[0,2]$ approximated using 100 equally spaced grid points. Number of Monte-Carlo replications: 1,000. Sample size: $M = 100$, $J_{m0} = 15$ for all m, $J_{m1} = 16$ for all m, $T = 2$.

with the non-linear GMM problem is given by:

$$\begin{aligned} \text{Structural equation:} \quad & \Delta\rho_j(\hat{\lambda}) = b_0 + b_1 \Delta J_{jm}(\hat{\lambda}) + \text{Error} \\ \text{First-stage:} \quad & \Delta J_{jm}(\hat{\lambda}) = \pi_0 + \pi_1 w^1_{jm} + \pi_2 w^2_{jm} + u_{jm} \end{aligned}$$

where $\Delta J_{jm} = \partial\sigma_j^{-1}(\boldsymbol{s}_{m1}, \boldsymbol{x}_{m1}; \hat{\lambda})/\partial\lambda - \partial\sigma_j^{-1}(\boldsymbol{s}_{m0}, \boldsymbol{x}_{m0}; \hat{\lambda})/\partial\lambda$ is the change in the Jacobian of the inverse-demand. Since w^1_{jm} is an indicator variable equal to one for products that are "closest" to the new entrant, π_1 measures the average change in the slope of the inverse demand for "treated" products, relative to products in the same market that are less exposed to the new entry (i.e. "control" group). Similarly, the coefficient π_2 allows the intensity of the treatment effect to vary with distance.

Figure 4 illustrates the ability of this identification strategy to eliminate the simultaneity bias associated with the endogenous location of products. The dash curves correspond to the Kernel density of the parameters estimated using the "difference-in-difference" moment conditions (4a), or the full "Differentiation IVs" moments (4b).[25]

The data generating process is designed so that the correlation between ξ_{jmt} and the Euclidian distance between rival products is -0.25. As Figure 4b illustrates, this leads to an attenuation

[25]The Differentiation IVs specification combines the sum of square of characteristic differences (i.e. quadratic IV), and the number of competing products within one standard-deviation (i.e. local IV).

bias in the estimate of the travel cost parameter obtained using standard instruments ($\hat{\lambda} \approx 1.89$, compared to $\lambda^0 = 4$). Since products located in "denser" regions of the product space have higher quality, the GMM specification that exploits variation in the distance to all products wrongly infer that consumers have a small disutility from distance. Figure 4b illustrates that the difference-in-difference moment conditions eliminate this bias. The distribution is centered around $\lambda^0 = 4$, and the average bias is less than 1% of the parameter value.

Comparing the two distributions, it is important to note that by exploiting solely the variation created by the entry of a new product, the difference-in-difference GMM estimator is less precise, and the distribution of $\hat{\lambda}$ is less well approximated by the normal density than the specification that uses the larger set of instruments. In Figure 4b the p-value associated with Shapiro-Wilk normal test is 11%, compared to less than 1% in Figure 4a. This suggests that the asymptotic approximation used to conduct inference on λ is less likely to be valid when the model is estimated solely using quasi-experimental variation; therefore requiring larger sample sizes or inference methods that are robust to weak identifications.

3.5 Comparison with the Optimal IV Approximation

Finally, we conclude this section by comparing the performance of the Differentiation IVs, with the approximation to the optimal IV proposed by Berry et al. (1999) and Reynaert and Verboven (2013).

Recall that, abstracting away from concerns related to heteroskedasticity, the instrument vector that minimizes the asymptotic variance of the parameter estimates is given by the conditional expectation of the jacobian of the residual function (Amemiya (1977), Chamberlain (1987)):

$$A_j^*(\boldsymbol{x}_t) \quad = \quad E\left[\frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|\boldsymbol{x}_t\right] = \left\{-\boldsymbol{x}_{jt}, E\left[\frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}\Big|\boldsymbol{x}_t\right]\right\}$$

This is very intuitive: Because the asymptotic distribution of $(\boldsymbol{\Sigma}, \boldsymbol{\beta})$ is derived from a first-order approximation of the residual function, the most efficient instruments correspond to the best-predictor of the slopes of that function with respect to each of the parameters.[26]

This theoretical lower-bound cannot be achieved in practice since the model is semi-parametric in ξ_{jt}. Rather than using non-parametric regression techniques to estimate $A_j^*(\boldsymbol{x}_t)$ (as in Newey (1990)), Berry et al. (1999) proposed the following heuristic approximation to the optimal IV:

$$E\left[\frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|\boldsymbol{x}_t\right] \approx \frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\xi_{jt}=0, \forall j,t} = \tilde{A}_j(\boldsymbol{x}_{jt}|\boldsymbol{\theta}). \tag{34}$$

Since the instrument vector depends on $\boldsymbol{\theta}$, users must first obtain an estimate of the parameters,

[26] See Newey (1993) for an illumination discussion.

denoted by $\boldsymbol{\theta}^1$. This leads to a two-step estimator: (i) estimate $\boldsymbol{\theta}^1$ by GMM using instrument vector $\boldsymbol{z}_{jt}$, and (ii) construct $\tilde{A}_j(\boldsymbol{x}_{jt}|\boldsymbol{\theta}^1)$ and estimate $\hat{\boldsymbol{\theta}}$ by GMM. The second step corresponds to a just-identified system of moment conditions.

When prices enter non-linearly in the model, a similar heuristic can be used to avoid taking an expectation over the second set of endogenous variables:

$$E\left[\frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \Big| \boldsymbol{x}_t, \boldsymbol{w}_t\right] \approx \frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}\Big|_{p_{jt}=\hat{p}_{jt}, \xi_{jt}=0, \forall j,t} = \tilde{A}_j(\boldsymbol{x}_{jt}|\boldsymbol{\theta}), \tag{35}$$

where $\hat{p}_{jt} \approx E(p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t)$ is a "reduced-form" model for prices independent of ξ_{jt}. Berry et al. (1999) proposed to compute $\hat{p}_{jt}$ by solving the equilibrium pricing game at the initial estimate $\boldsymbol{\theta}^1$, using only observed/pre-determined variables. Reynaert and Verboven (2013) uses linear regressions and the existence of a price instrument to compute $\hat{p}_{jt}$. As discussed above, we follow this latter approach in our simulations.

Reynaert and Verboven (2013) conducted a series of Monte-Carlo simulations to illustrate that this heuristic approximation to the optimal IV leads to substantial efficiency gains over the standard instruments proposed in Berry et al. (1995) (i.e. sum of rival characteristics). One remaining question however is to what extent the approximation remains valid when the first-stage estimates are not consistent, which is the case for instance with weak instruments. To illustrate when consistency is likely to matter, we first study two simple mixed-logit models: (i) normal random-coefficient, and (ii) Hotelling. These two models satisfy our "linear-in-characteristic" random-coefficient assumption and have the following indirect-utility function:

$$\begin{aligned} \text{Normal RC:} \qquad & u_{ijt} = \delta_{jt} + \lambda \eta_i x_{jt}^{(2)} + \epsilon_{ijt} \\ \text{Hotelling:} \qquad & u_{ijt} = \delta_{jt} - \lambda \left(\eta_i - x_{jt}^{(2)}\right)^2 + \epsilon_{ijt}. \end{aligned}$$

where $\eta_i \sim \mathcal{N}(0,1)$ and $x_{jt}^{(2)} \sim \mathcal{N}(0,1)$.[27] For our purpose, the key distinction between these two models is that the value of λ in the "Normal RC" model only affects the magnitude of the elasticity of substitution, and **not** the relative ranking of each products' cross-elasticty (which is function only of x's). In contrast, in the Hotelling model, when λ goes from positive to negative, the identity of the "closest" competitor changes from the "closest" x to the "furthest" x. We use this stark distinction between the two models to illustrate how consistency of the first-stage, λ^1, affects the performance of the heuristic approximation.

Table 7 summarizes the results of 1,000 Monte-Carlo replication simulations. The first five rows correspond to different values of the initial parameter used to evaluate the jacobian. In both

[27] The Hotelling mixed-logit model is linear in characteristic: $u_{ijt} = \delta_{jt} - \lambda \left(\eta_i - x_{jt}^{(2)}\right)^2 + \epsilon_{ijt} = \delta_{jt} - \lambda \eta_i^2 + \lambda \left(x_{jt}^{(2)}\right)^2 - 2\lambda \eta_i x_{jt}^{(2)} + \epsilon_{ijt}$

Table 7: Optimal IV approximation with alternative initial parameter values

	Normal RC			Hotelling		
	λ^1	bias	rmse	λ^1	bias	rmse
Optimal IV approx.:						
(1)	0.5	0.001	0.027	4	-0.003	0.140
(2)	1.5	0.001	0.026	**2**	**-0.004**	**0.126**
(3)	2	**0.001**	**0.026**	0	-0.079	0.509
(3)	2.5	0.001	0.026	-1	-0.344	1.687
(4)	3	0.002	0.028	-2	-0.282	1.254
Differentiation IV	—	0.001	0.031	—	0.017	0.310

specifications the true value of parameter is $\lambda^0 = 2$. The numbers in bold correspond to GMM results obtained by setting the first-stage parameter equal to the true parameter value . The rest of the rows correspond to different levels of inconsistencies. For the "Normal RC" mode, we consider a grid between 0.5 and 3. For the Hotelling model, we consider grid between -2 (wrong sign) and 4.

Looking first at the "Normal RC" model, the performance of the optimal IV approximation estimator is remarkably robust to inconsistencies in the first-stage parameter values. The efficiency gains from using the "true" parameter value are fairly small (i.e. 0.026 vs 0.028). This is consistent with the results presented in Reynaert and Verboven (2013) who focus only on models with multiplicative random-coefficients.

The results from the "Hotelling" specification are very different. The first two rows show that using using an inconsistent first-stage parameter with the correct sign does not reduce dramatically the precision of the estimates (i.e. 0.14 vs 0.126). However, using first-stage values that are inconsistent and have the wrong sign lead to large attenuation biases, and very imprecise estimates. The RMSE in the last two rows are more than 10 times larger than in specification (2) (i.e. true σ_1). This suggests that the consistency of the first-stage estimate is important for the validity of the heuristic approximation approach, especially when the substitution patterns depend on the sign of the parameter values.

The last row of Table 7 reports the results obtained with the Differentiation IVs. To obtain these results we combine the sum of square of characteristic differences, and the number of local competitors. When using unbiased first-stage parameter, the optimal IV approximation improves the precision of the estimates 60% in the Hotelling model, and by 17% in the Normal RC model. However, these efficiency gains are quickly eliminated when the first-stage parameter is set far from $\boldsymbol{\theta}^0$. This is an important advantage of the Differentiation IVs, since their exact structure does not depend on the availability of consistent estimates, or on prior the knowledge of the model of differentiation (e.g. hotelling versus normal).

The previous example is very stylized. Another setting in which the sign and magnitude of

Table 8: Monte-Carlo simulation results for correlated random-coefficient specification with optimal IV approximation and inconsistent initial parameter values

Choleski matrix	True	Opt. IV: $\boldsymbol{\theta}^1 \sim N(0,1)$ bias	rmse	se	Opt. IV: $\boldsymbol{\theta}^1 \sim N(0,4)$ bias	rmse	se	Diff. IV: Quadratic bias	rmse	se
	(1)	(2)	(3)	(4)	(4)	(5)	(6)	(7)	(8)	(9)
$\log c_{11}$	0.69	0.00	0.22	5.42	0.01	1.22	11.92	-0.00	0.03	0.03
$\log c_{22}$	0.55	-0.01	0.19	2.50	-0.16	2.36	192.70	-0.00	0.04	0.04
$\log c_{33}$	0.49	-0.02	0.15	0.46	-0.44	2.69	++	-0.00	0.04	0.04
$\log c_{44}$	0.46	-0.22	1.83	++	-1.78	5.57	++	-0.00	0.04	0.04
c_{21}	-1.00	0.01	0.47	4.51	0.03	0.77	781.85	0.00	0.06	0.06
c_{31}	1.00	0.00	0.33	0.86	-0.02	0.63	23.48	-0.00	0.07	0.07
c_{32}	-0.58	0.02	0.27	2.69	0.03	0.56	285.80	0.00	0.07	0.08
c_{41}	1.00	0.00	0.23	1.37	0.00	0.58	333.93	0.00	0.07	0.07
c_{42}	-0.58	0.01	0.23	2.69	0.04	0.50	484.88	0.00	0.08	0.08
c_{43}	0.41	0.00	0.23	1.59	0.03	0.52	++	0.00	0.08	0.08

$\boldsymbol{\theta}$ determines substitution patterns is the correlated random-coefficient model studied in Section 3.2. To illustrate the importance of using consistent estimates in the first-stage, we implement the optimal IV approximation using pseudo-random random values that are not centered around the truth. The results are summarized in Table 8. In columns (2)-(4), each element of $\boldsymbol{\theta}^1$ is drawn from a standard-normal distribution, while columns (4)-(6) they are drawn from a normal distribution with a standard-deviation of 2. The last four rows reproduce the results obtained using the Differentiation IVs discussed above.

The results are in line with single-dimension example. Using inconsistent parameter estimates to approximate the optimal instruments lead to a weak identification problem, associated with very noisy and often biased parameter estimates. In addition, as we increase the variance of $\boldsymbol{\theta}^1$, the precision and bias of $\hat{\boldsymbol{\theta}}$ both increase substantially. The contrast with the Differentiation IVs is quite striking: the average RMSEs are roughly 5 times smaller with the Differentiation IVs than with the less noisy optimal IV approximation.

Overall, these results suggest that using consistent first-stage estimates is important in order for the optimal IV approximation to perform well, and therefore that using strong first-stage instruments is crucial to estimate $\boldsymbol{\theta}^1$. For instance, a valid strategy to improve the efficiency of the estimates is to obtain first-stage estimates using the instruments proposed in this paper, and then construct an approximation to the optimal IV. The second-stage can be conducted using the heuristic approximation proposed by Berry et al. (1999) or Reynaert and Verboven (2013), or using non-parametric regressions as discussed in Newey (1993).

We illustrate this two-step approach using the model with endogenous prices studied in Section 3.3. Table 9 summarizes the results. The top-panel corresponds to the GMM estimates obtained using three alternative Differentiation IV: (i) local competition, (ii) sum of square of characteristic

Table 9: Monte-Carlo simulation results for endogenous price specification and optimal IV approximation

			Diff. IV = Local			Diff. IV = Quadratic			Diff. IV = Sum		
		True	bias	se	rmse	bias	se	rmse	bias	se	rmse
1st-stage	λ_p	-4	0.02	0.27	0.28	0.02	0.53	0.55	1.01	2.66	2.09
	β_0	50	-0.26	3.92	3.92	-0.28	7.36	7.45	-9.63	26.48	20.46
	β_x	2	-0.02	0.46	0.45	-0.02	0.47	0.47	0.34	1.11	0.83
	β_p	-0.2	0.01	0.37	0.37	0.01	0.31	0.32	-0.66	1.76	1.37
2nd-stage	λ_p	-4	0.00	0.24	0.23	0.00	0.24	0.23	0.01	0.26	0.31
	β_0	50	-0.07	3.99	3.84	-0.06	3.72	3.65	0.05	4.32	4.61
	β_x	2	-0.01	0.48	0.47	-0.01	0.41	0.41	0.03	0.52	0.51
	β_p	-0.2	0.01	0.36	0.36	0.00	0.31	0.32	-0.03	0.40	0.40

differences, and (iii) sum of rival characteristics. In each specification we use the residual cost-shock, ω_{jt}, as a price instrument. In the bottom-panel, we use the GMM results from the corresponding specification to construct an approximation to the optimal IV, as described in equation (34). Each entry is averaged over 1,000 Monte-Carlo replications.

The results suggest that for the most part the Berry et al. (1999) approximation successfully correct the weak identification problem. For instance, the sum of rival characteristics specification is associated with very noisy estimates of λ_p in the top panel, but the average bias and RMSE are mostly comparable across columns in the bottom panel. Similarly, the RMSE of λ_p estimated with the quadratic Differentiation IVs is roughly 50% smaller in the second-stage. The efficiency gains are much smaller the first specification, mostly because the local Differentiation IVs is already a very strong instrument (see Table 6b).

Across columns, we see that using stronger instruments in the first stage lead to more precise results in the second stage (i.e. RMSE = 0.31 vs 0.23). This should be thought of as a lower bound on the efficiency gains of using strong versus weak first-stage instruments. As we saw in the "Hotelling" vs "Normal RC" examples above (see Table 7), the multiplicative random-coefficient model is not very sensitive to the value first-stage parameter, and the efficiency loss from using inconsistent initial parameter values is small. Also, Reynaert and Verboven (2013)'s simulation results suggest that the heuristic approximation is becoming weaker as the the number of random-coefficients increases beyond four. Since we consider only a single-dimension example, it is likely that the efficiency gain from using stronger first-stage instruments would be larger as the number of random-coefficients increases. Importantly, this is not the case for the Differentiation IVs that we proposed (see Tables 3 and 5).

A Proof of Propositions 1 and 2

A.1 First Proposition

Proposition 1 can be restated as follows.

Proposition 3. *In the linear characteristics model the market inverse function can be expressed as*

$$D_j^{-1}(s_0, s_1, \ldots, s_J; X) \quad = \quad G\left(s_j, \{s_k, d_{jk}\}_{k \neq j}\right) + C$$

where $d_{jk} = x_k - x_j$ *and* C *is a constant that is common to all products* $j = 1, \ldots, J$.

The proposition implies that all the cross sectional variation in the inverse function comes from the component

$$G\left(s_j, \{s_k, d_{jk}\}_{k \neq j}\right) = G\left(s_j, F_j(s, d)\right)$$

where we have equivalently expressed the second argument as the empirical distribution of (s_k, d_{jk}) among products $k \neq j$ (which includes the outside good 0 in this sample). It is important to note that from this empirical distribution, we can only recover the set of the differences d_{jk} but cannot isolate the difference with respect to any particular product, and also cannot recover x_j itself from this distribution (because we cannot identify the outside good in this set). This brings to light that the cross sectional variation in the inverse function does not actually depend on a product's level of own x_j, but rather the distribution of differences d_{jk} for $k \neq j$ this product faces.

We will spend the rest of this section proving the result.

Step 1

The first step is to re-parameterize the demand function $D_j(\delta_1, \ldots, \delta_J)$ in terms of

$$t_j = \frac{exp(\delta_j)}{\sum_{l=0}^{J} \exp(\delta_l)}.$$

The advantage of this re-parameterization is that it is an alternative location normalization (requiring that all products t's to sum to one) that does not create an asymmetry between the outside good 0 and the inside goods $j > 1$. This will be analytically more convenient than the standard normalization of $\delta_0 = 0$. But they are mathematically identical. In particular observe that

$$T_j = \log(t_j) = \delta_j + C$$

where C is a constant that is common to all products in a market (that can be solved by recognizing $\log t_0 = -C$).

We can thus express demand in terms of this re-parameterization, i.e.,

$$u_{ij} = u\left(t_j, x_j, \theta_i\right) = T_j + \sum_{k=1}^{K} \sigma_k v_{ik} x_{jk} + \epsilon_{ij}$$

and $\mathcal{D}_j\left(t_0, \ldots, t_J\right) =$

$$\int \mathbf{1}\left[u\left(t_j, x_j, \theta\right) \geq u\left(t_k, x_k, \theta\right) \ \forall k = 0, \ldots, J, \ k \neq j\right] \ dF\left(\theta\right). \tag{36}$$

We then have that

$$\mathcal{D}_j\left(t_0, \ldots, t_J\right) = D_j\left(\delta_1, \ldots, \delta_J\right).$$

(because an additive constant does not change preferences). Moreover we have that

$$\log \mathcal{D}_j^{-1}\left(s_0, \ldots, s_J\right) + C = D_j^{-1}\left(s_0, \ldots, s_J\right).$$

Our strategy moving forward is to show that

$$\mathcal{D}_j^{-1}\left(s_0, \ldots, s_J\right) = \mathcal{D}^{-1}\left(s_j, \left\{s_k, d_{jk}\right\}_{k \neq j}\right). \tag{37}$$

Then defining $G = \log D^{-1}$ will give us the Theorem.

Step 2

We now establish 3 properties of $\mathcal{D}_j\left(t_1, \ldots, t_J\right)$: symmetry, anonymity, and translation invariance. Each of these properties will then be preserved by the inverse mapping $\mathcal{D}_j^{-1}$. To establish these properties let us define a product j's state ω_j as

$$\omega_j = \left(t_j, x_j\right)$$

and note that

$$\mathcal{D}_j\left(t_0, \ldots, t_J\right) = \mathcal{D}_j\left(\omega_j, \omega_{-j}\right).$$

The following two properties are relatively straightforward to show using the definition of demand (36) and the symmetry of the idiosyncratic errors (**??**). The first property is

Definition 1. *The function $\mathcal{D}_j\left(\omega_j, \omega_{-j}\right)$ is* symmetric *if $\mathcal{D}_j\left(\omega_j, \omega_{-j}\right) = \mathcal{D}_k\left(\omega_j, \omega_{-j}\right)$ for any $k \neq j$.*

This implies we can write $\mathcal{D}_j\left(\omega_j, \omega_{-j}\right) = \mathcal{D}\left(\omega_j, \omega_{-j}\right)$.

Definition 2. *The function $\mathcal{D}\left(\omega_j, \omega_{-j}\right)$ is anonymous if $\mathcal{D}\left(\omega_j, \omega_{\rho(-j)}\right)$ where ρ is any permutation of the indices $-j$.*

We note that symmetry and anonymity are the same properties that Doraszelski and Pakes (2007) use to reduce the dimensionality of value functions in dynamic games. These properties can be established for the demand functions $\mathcal{D}_j$.

There is one last property of demand we will exploit which is the following:

Definition 3. *The function $\mathcal{D}(\omega_j, \omega_{-j})$ is translation invariant if for any $c \in \mathbb{R}^K$ we have that*

$$\mathcal{D}\left(\omega_j + (0,c), \omega_{-j} + (\vec{0,c})\right) = \mathcal{D}(\omega_j, \omega_{-j})$$

where $(\vec{0,c})$ is the J dimensional vector consisting of elements $(0,c)$.

This property can be established using the linearity of the characteristics utility u_{ij} in x_j. It is important to note that the second argument in $\mathcal{D}$ includes the outside good.

Step 3

Now define the relevant state for the inverse mapping as

$$m_j = (s_j, x_j).$$

Then

$$\mathcal{D}_j^{-1}(s_0, \ldots, s_J) = \mathcal{D}_j^{-1}(m_j, m_{-j}).$$

Using the above properties of the demand function $\mathcal{D}$, we can establish precisely the same properties for $\mathcal{D}_j^{-1}$, namely symmetry, anonymity, and translation invariance. Thus we have that

$$\begin{aligned} \mathcal{D}_j^{-1}(m_j, m_{-j}) &= \mathcal{D}^{-1}\left(m_j + (0, -x_j), m_{-j} + (0, \vec{-x_j})\right) \\ &= \mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k \neq j}\right) \end{aligned}$$

where the first equality follows from symmetry and translation invariance, and the second equality follows from anonymity. We have thus succeeded in establishing (37) and hence Theorem 3.

A.2 Second Proposition

Let $X = (x_0, \ldots, x_J)$ be the entire market menu of product characteristics. We assume here for simplicity that X is fully independent of $\xi = \xi_1, \ldots, \xi_J$. Consistent with the symmetry of the model, the distribution F_ξ is assumed to have a symmetric distribution. Then we have the following result which suffices to establish Proposition 2 in the paper.

Proposition 4. *The conditional expectation of interest in the model can be expressed as*

$$\begin{aligned} E\left[\mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k\neq j}\right) \mid X\right] &= E\left[\mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k\neq j}\right) \mid \{d_{jk}\}_{k\neq j}\right] \\ &= E\left[\mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k\neq j}\right) \mid F_j(d)\right] \end{aligned}$$

where $F_j(d)$ is the empirical distribution of the sample of differences $\{d_{jk}\}_{k\neq j}$.

We will only sketch here the main ideas of the proof (with details to be filled later). Assume that the d_{jk} can be canonically ordered (based on some complete ordering in $\mathbb{R}^K$, such as the lexicographic ordering) such that $\tilde{d}_{j1} \leq \cdots \leq \tilde{d}_{jK}$ where $\tilde{d}_{jl}$ is the lth largest from the $\{d_{jk}\}_{k\neq k}$. Then we can express

$$\mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k\neq j}\right) = D^{-1}\left(\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ}; \tilde{d}_{j1}, \ldots, \tilde{d}_{jJ}\right)$$

where $\tilde{s}_{j0}$ is s_j and $\tilde{s}_{ji}$ is the market share corresponding to the product with difference $\tilde{d}_{ji}$. Now it can be shown that the distribution

$$F_{\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ} \mid X} = F_{\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ} \mid \tilde{d}_{j1}, \ldots, \tilde{d}_{jJ}}.$$

That is $\tilde{d}_{j1}, \ldots, \tilde{d}_{jJ}$ is a sufficient statistic of the market menu X to determine the distribution of the shares $(\tilde{s}_{j0}, \ldots, \tilde{s}_{jJ})$. We then have that

$$\begin{aligned} E\left[\mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k\neq j}\right) \mid X\right] &= E\left[D^{-1}\left(\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ}; \tilde{d}_{j1}, \ldots, \tilde{d}_{jJ}\right) \mid X\right] \\ &= E\left[D^{-1}\left(\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ}; \tilde{d}_{j1}, \ldots, \tilde{d}_{jJ}\right) \mid \tilde{d}_{j1}, \ldots, \tilde{d}_{jJ}\right] \\ &= E\left[\mathcal{D}^{-1}\left(s_j, \{(s_j, d_{jk})\}_{k\neq j}\right) \mid F_j(d)\right] \end{aligned}$$

References

Ai, C. and X. Chen (2003, November). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

Amemiya, T. (1974). The non-linear two-stage least-square estimator. *Journal of Econometrics 2*, 105–110.

Amemiya, T. (1977). The maximum-likelihood and non-linear three-stage least-squares estimator in the general nonlinear simultaneous equation model. *Econometrica 45*, 955–968.

Angrist, J. and J.-S. Pischke (2009). *Mostly harmless econometrics: An Empiricist's companion.* Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010, Spring). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Armstrong, T. (2014, August). Large market asymptotics for differentiated product demand estimators with economic models of supply. working paper, Yale University.

Ashenfelter, O., D. Hosken, and M. Weinberg (2009). Generating evidence to guide merger enforcement. CEPS working paper No. 183.

Berry, S. (1994). Estimating discrete choice models of product differentiation. *Rand Journal of Economics 25*, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review 89*(3), 400–430.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*, 1–68.

Berry, S. T., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica 81*(5), 2087–2111.

Berry, S. T. and P. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1798.

Bresnahan, T. F. (1981). Departures from marginal-cost pricing in the american automobile industry. *Journal of Econometrics 17*(2), 201.

Bresnahan, T. F. (1987). Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics 35*(4, The Empirical Renaissance in Industrial Economics), 457–482.

Bresnahan, T. F., S. Stern, and M. Trajtenberg (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics 28*(No. 0, Special Issue in Honor of Richard E. Quandt), S17–S44.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics 34*(305—334).

Ciliberto, F., C. Murry, and E. Tamer (2016, May). Market structure and competition in airline markets. working paper, University of Virginia.

Conlon, C. (2013, September). The empirical likelihood mpec approach to demand estimation. working paper, Columbia University.

Conlon, C. and J. H. Mortimer (2015, April). Efficiency and foreclosure effects of vertical rebates: Empirical evidence. working paper, Columbia University.

Cragg, J. G. and S. G. Donald (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory 9*, 222–240.

Davis, P. (2006, Winter). Spatial competition in retail markets: Movie theaters. *Rand Journal of Economics*.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics 117*(1), 55–93.

Donald, S. G., G. W. Imbens, and W. K. Newey (2008). Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics 152*, 28–36.

Doraszelski, U. and A. Pakes (2007). A framework for applied dynamic analysis in io. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*, Volume 3, Chapter 30, pp. 1557–2440. Elsevier.

Eizenberg, A. (2014). Upstream innovation and product variety in the u.s. home pc market. *Review of Economic Studies*.

Fan, Y. (2013, November). Ownership consolidation and product characteristics: A study of the u.s. daily newspaper market. *American Economic Review 103*(5), 1598–1628.

Farias, V., D. Saure, and G. Y. Weintraub (2012, Summer). An approximate dynamic programming approach to solving dynamic oligopoly models. *Rand Journal of Economics 43*(2), 253–282.

Gentzkow, M. and J. Shapiro (2010). What drives media slant? evidence from u.s. news- papers. *Econometrica 78*, 35–71.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5).

Hastings, J., T. Kaine, and D. Staiger (2009, May). Heterogeneous preferences and the efficacy of public school choice.

Hausman, J. A. and D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica 52*(5), 1219–1240.

Houde, J.-F. (2012, August). Spatial differentiation and vertical mergers in retail markets for gasoline. *American Economic Review 102*(5), 2147–2182.

Jorgensen, D. W. and J.-J. Laffont (1974). *Efficient estimation of nonlinear simultaneous equations with additive disturbances*, Volume 3 of *Annals of Economic and Social Measurement*, pp. 615–640. NBER.

Metaxoglou, K. and C. R. Knittel (2014, October). Estimation of random coefficient demand models: Two empirisists' perspectives. *The Review of Economic and Statistics 96*(1).

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica 58*(809-837).

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In G. S. Maddala, C. R. Rao, and H. D. Vinod (Eds.), *Handbook of Statistics*, Volume 11. Elsevier.

Nielson, C. (2013, November). Targeted vouchers, competition among schools, and the academic achievement of poor students.

Pakes, A. (1994). Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interactions. In J.-J. Laffont and C. A. Sims (Eds.), *Advances in Econometrics: The sixth world congress of the econometric society*, Volume 2. Cambridge University Press.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy 110*, 705–729.

Pinkse, J., M. E. Slade, and C. Brett (2002). Spatial price competition: A semiparametric approach. *Econometrica 70*(3), 1111–1153.

Reynaert, M. and F. Verboven (2013). Improving the performance of random coefficients demand models: The role of optimal instruments. *Journal of Econometrics 179*(1), 83–98.

Romeo, C. J. (2010). Filling out the instrument set in mixed logit demand systems for aggregate data. working paper, US Department of Justice.

Sanderson, E. and F. Windmeijer (2016). A weak instrument f -test in linear iv models with multiple endogenous variables. *Journal of Econometrics 190*, 212–221.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*, 518 – 529.

Stock, J. H. and M. Yogo (2005). *Testing for weak instruments in linear IV regression*, pp. 80–108. Cambridge University Press.

Thomadsen, R. (2007). The effect of ownership structure on prices in geographically differentiated industries. *Rand Journal of Economics*.

Verboven, F. (1996). International price discrimination in the european car market. *Rand Journal of Economics 27*(2), 240–268.

Wright, J. H. (2003). Detecting lack of identification in gmm. *Econometric Theory 19*(2), 322–330.