Measuring Substitution Patterns in Differentiated Products Industries*

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

[1]University of Wisconsin-Madison
[2]Wharton School, University of Pennsylvania

January 26, 2016

Introduction

A very 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. Thus substitution patterns are the key to 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 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

*Preliminary and Incomplete. Please do not cite without the permission of the authors. Gandhi and Houde acknowledge financial support from the NSF.

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 restrictive special cases of the model such as nested logit.

This basic 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. Although the parameters that govern substitution patterns in the mixed logit demand system are in principle identified under a conditional moment restriction between unobserved and observed product characteristics (as formally shown recently by Berry and Haile (2014)), estimation of these parameters in finite samples requires constructing a finite set of instruments from the conditional moment restriction. An instruments 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 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.

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 these optimal instruments (as established by e.g., Newey (1990), Newey (1993), 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 optimal instruments (linear or quadratic) will lead to basis function that grow exponentially large

in number with the number of products in the market. This curse of dimensionality has made it hard if not impossible to apply basic insights from the optimal IV literature to demand estimation in markets with more than a few products for the typical size data set.

Our main result in the paper is that we show that this curse of dimensionality can be solved by using the implicit restrictions that the demand structure of the model places on the form of the optimal instruments. In particular we show that the symmetry properties implicit in the demand model imply that the optimal 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. The basis functions themselves are intuitive - 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]

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

[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.

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 are indeed complementary to one another in an empirically powerful way.

1 Identification of the random-coefficient model

1.1 Description of the Model

We briefly review the *random coefficients* utility model that is widely used as a foundation for differentiated product demand. Our presentation of the model largely follows the setup in Berry, Levinsohn, and Pakes (1995). The key difference is that we intentionally exclude an endogenous price from the model in order to isolate the problem of identifying and estimating substitution patterns. Endogenous prices do not fundamentally change the main results of our analysis, and we discuss the case with endogenous product attributes in Section 2.2.

Consider a market t with $n_t + 1$ 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 $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{n_t,t})$ as a summary of *market structure* - the entire menu of product characteristics available to consumers in market t (i.e. $n_t \times K$ matrix).

The data-set contains a panel of shares and product characteristics, i.e.,

$$\text{Data} = \left\{ \{s_{jt}, \boldsymbol{x}_{jt}\}_{j=1}^{n_t} \right\}_{t=1}^{T}. \tag{1}$$

Let $n = \sum_{t=1}^{T} n_t$ denotes the sample size.

Following, Berry and Haile (2014), the product characteristics space is partitioned into two subsets: $\boldsymbol{x}_{jt} = \left\{ \boldsymbol{x}_{jt}^{(1)}, \boldsymbol{x}_{jt}^{(2)} \right\}$ where we refer to $\boldsymbol{x}_{jt}^{(1)}$ as the *regular* product characteristics, and $\boldsymbol{x}_{jt}^{(1)}$ as the *special* product characteristics (these names reflecting the roles they play in the identification argument in Berry and Haile (2014)). The regular characteristics, $\boldsymbol{x}_{jt}^{(1)}$, are product attributes for which consumers have homogenous preferences. In contrast, consumers have heterogenous tastes for the special characteristics $\boldsymbol{x}_{jt}^{(2)}$. Le $K_2 = \dim\left(\boldsymbol{x}_{jt}^{(2)}\right)$ be the dimension of the special characteristic.

We assume that consumers have linear preferences for product characteristics:

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^{K_2} x_{jt,k}^{(2)} v_{jk} + \varepsilon_{jt}, \tag{2}$$

where $\delta_{jt} = \boldsymbol{\beta}\boldsymbol{x}_{jt} + \xi_{jt}$ is typically labelled the "mean utility" of product j. According to this specification, the taste of consumer i for characteristic k is denoted by $\beta_{ik} = \beta_k + v_{ik}$ where v_{ik} is a mean zero taste shock for characteristic k and ϵ_{ijt} is an idiosyncratic taste for product j in the underlying population of consumers. Given the linearity of consumer utility we can normalize the characteristics of an outside good 0 such that $\boldsymbol{x}_{0t} = 0$ and $\xi_{0t} = 0$.[4]

Consumers have heterogeneous tastes in the population. The standard approach in the literature assumes this heterogeneity takes a mixed-logit form which is our focus here, although our analysis below applies to a much more general class of random utility models. Thus we have that the idiosyncratic taste is distributed $\epsilon_{ij} \overset{iid}{\sim} \text{T1EV}(0,1)$. The idiosyncratic components of random coefficients, $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iK_2})$, are IID random variables with parametric probability distribution function $f(\boldsymbol{v}_i; \Sigma)$. Let $\boldsymbol{\theta}^0 = \left(\boldsymbol{\beta}^0, \boldsymbol{\Sigma}^0\right)$ denotes the true parameters determining the distribution of consumer tastes.

Importantly, we assume that the distribution of random-coefficients is common across markets. We do so in order to focus on the variation generated by having different menus of product characteristics across markets. In Section 2.2, we discuss how to adapt our results to the case in which market-specific consumer characteristics are the main source of variation (e.g. Nevo (2001)).

If each consumer i chooses the product $j \in \{0, \ldots, n_t\}$ that maximizes his/er utility, then we can integrate over the distribution of consumer choices to yield a market share for each product j that is given by

$$s_{jt} = \sigma_j\left(\boldsymbol{x}_t^{(2)}, \boldsymbol{\delta}_t; \boldsymbol{\Sigma}\right) \quad = \quad \int \frac{\exp\left(\sum_k v_{ik} x_{jt,k}^{(2)} + \delta_{jt}\right)}{1 + \sum_{j'=1}^{n_t} \exp\left(\sum_k v_{ik} x_{j't,k}^{(2)} + \delta_{j't}\right)} f(\boldsymbol{v}_i; \boldsymbol{\Sigma}) d\boldsymbol{v}_i \tag{3}$$

where $\boldsymbol{x}_t^{(2)} = \left(\boldsymbol{x}_{1t}^{(2)}, \ldots, \boldsymbol{x}_{n_t,t}^{(2)}\right)$ and $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{n_t,t})$, and s_{jt} is the observed market-share of (j,t). We rule out the possibility that observed market shares are measured with errors.

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

1.2 Identification

An identification problem arises because for any hypothetical random-coefficient distribution there exists a unique mean utilities for products $j = 1, \ldots, n_t$ that rationalizes the market shares, i.e.,

$$\begin{aligned} s_{jt} &= \sigma_j\left(\boldsymbol{x}_t^{(2)}, \boldsymbol{\delta}_t; \boldsymbol{\Sigma}\right) \quad j = 1, \ldots, n_t \\ &\iff \xi_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) = \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\Sigma}\right) - \boldsymbol{x}_{jt}\boldsymbol{\beta} \quad j = 1, \ldots, n_t \end{aligned} \tag{4}$$

where the inverse demand σ_j^{-1} is uniquely defined following Berry, Levinsohn, and Pakes (1995) and Berry, Gandhi, and Haile (2013). The mapping $\xi_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right)$ corresponds to the structural residual of the model. When evaluated at the true distribution parameters $\boldsymbol{\theta}^0$, it is equal to true unobserved product attributes: $\xi_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right) = \xi_{jt}$. Observe that the additive structure of the model implies that inverse demand σ_j^{-1} is only function of the parameters that determine the heterogeneity of taste differences ($\boldsymbol{\Sigma}$), and not of the mean valuation parameters ($\boldsymbol{\beta}$). As a result we will refer to the parameters in $\boldsymbol{\Sigma}$ as the non-linear parameters because they enter the residual function $\xi_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right)$ in a non-linear way (through σ_j^{-1}).

To identify the model we follow Berry, Levinsohn, and Pakes (1995), and employ a restriction with a long history in empirical industrial organization: we assume that the unobserved quality ξ_{jt} is independent of the market structure $\boldsymbol{X}_t$, i.e.,

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

Although it has a variety of economic motivations (such as common view in the literature that non-price product attributes are fixed in the short run and are costly to adjust), our purpose here is not to justify the assumption, but rather to extract its empirical usefulness. If we apply the conditional moment restriction to both sides of the identification condition:

$$E\left[\xi_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right) \mid \boldsymbol{x}_t\right] = 0. \tag{5}$$

The model is identified if θ^0 is the unique parameter vector that solves the continuum of moment restrictions (5). Berry and Haile (2014) have recently shown primitive conditions under which a unique θ^0 solves these moments and hence the model parameters are identified. In what follows we assume that those conditions are met, and that the model is identified the CMR.

In order to use this source of identification in estimation with finite samples (such as our data (1)), we must translate the continuum of moment restrictions into a finite set of restrictions. We can exploit the fact that the CMR implies that $\xi_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right)$ is uncorrelated with any function of the menu of product characteristics (see e.g., Newey (1993)). In particular, if we let $\boldsymbol{A}_j(\boldsymbol{x}_t)$ denote

a vector of L instrumental Variables (IVs), then it follows that

$$E\left[\boldsymbol{A}_j(\boldsymbol{x}_t)\xi_j\left(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\theta}^0\right)\right]=0$$

which constitutes a $1\times L$ vector of population moment conditions that we can take to estimation.

The fundamental empirical challenge is how to construct the instruments $\boldsymbol{A}_j(\boldsymbol{x}_t)$. In particular for each element of the parameter vector θ we need to include at least one instrument in $\boldsymbol{A}_j(\boldsymbol{x}_t)$. How should such 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 $\boldsymbol{\Sigma}$ the choice is far less obvious. Moreover these are the parameters that govern consumer heterogeneity and hence substitution patterns among product, and hence are critical from a policy perspective. As we show in Section 4, seemingly intuitive choices for the elements of $\boldsymbol{A}_j(\boldsymbol{x}_t)$ can very readily lead to a weak IV problem for the parameters $\boldsymbol{\Sigma}$.

To avoid using weak IV's, we can look to the optimal instruments for the non-linear parameters $\boldsymbol{\Sigma}$. The optimal instruments for the non-linear parameters correspond to the *best-predictor* of the Jacobian of $\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\Sigma}\right)$ with respect to $\boldsymbol{\Sigma}$ (see Jorgensen and Laffont (1974), Amemiya (1977), and Chamberlain (1987)):

$$D_{j,k}(\boldsymbol{x}_t)=E\left[\left.\frac{\partial\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)}|\boldsymbol{\Sigma}^0\right)}{\partial\boldsymbol{\Sigma}_k}\right|\boldsymbol{x}_t\right],\quad\forall k=1,\ldots,\dim\left(\boldsymbol{\Sigma}\right). \tag{6}$$

We will refer to $D_{j,k}(\boldsymbol{x}_t)$ as the reduced-form of the model implied by the CMR (i.e. the "optimal" fist-stage regression). Strong instruments $\boldsymbol{A}_j(\boldsymbol{x}_t)$ will be closely correlated to these optimal instruments. On the other hand if $\boldsymbol{A}_j(\boldsymbol{x}_t)$ is only *weakly* correlated with the slope of the residual function, standard asymptotic theoretical results break down, and $\hat{\theta}^{gmm}$ is not consistent and has a non-standard asymptotic distribution (see Stock and Wright (2000)).

There are two basic problems with applying the insight of optimal IV to the practical construction of instruments. The first is that the optimal instrument depends on the true value $\boldsymbol{\Sigma^0}$ which is unknown. Although one can plug consistent estimates of Σ^0 into the formula, obtaining these consistent estimates requires a choice of instruments and weak instruments in the first stage will not lead to good performance of the resulting approximation. This problem has been solved in the optimal IV literature following Newey (1990), and Donald et al. (2003). In particular if the collection of L instruments $\boldsymbol{A}_j(\boldsymbol{x}_t)$ can arbitrarily well approximate any function $a\left(\boldsymbol{x_t}\right)$, i.e.,

$$a\left(\boldsymbol{x}_t\right)\approx\boldsymbol{A}_j(\boldsymbol{x}_t)'\gamma_L$$

where the error of the approximation grows small as $L\to\infty$, then using $\boldsymbol{A}_j(\boldsymbol{x}_t)$ as instruments

(and letting L grow large with sample size) will replicate the asymptotic performance of the optimal instruments. This insight directly suggests the use of the basis functions from a low order polynomial approximations to $D_{j,k}(\boldsymbol{x}_t)$ as instruments.

Unfortunately even with this basic result in hand, approaching the optimal instrument is still very challenging in this environment. This is because the reduced-form of the model is a product-specific function of the entire menu of product characteristics in the market. Without further restrictions, non-parametrically estimating equation (6) is feasible **only** when the number of products is small, relative to the number of markets.[5] As a result any basis approximation to the function (6) suffers from a curse of dimensionality: the number of polynomial basis terms to approximate $D_{j,k}(\boldsymbol{x}_t)$ in any fixed order approximation grows exponentially with the number of products per market. To see this, note that for each product j, the non-parametric regression tracks the variation in its relative market across different market structures $\boldsymbol{x}_t$, which can be a very high dimensional vector. Even for moderately small markets with ten products and five characteristics, the number of independent markets required is several times larger than what is conceivably available to researchers. When the number of products grows with the number of markets, the reduced-form of the model is not identified. Intuitively, changes in market structure, from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t'}$, affect each products differentially due to market segmentation. Therefore the expectation of the inverse-demand is a product-specific function.

Our main result shows how the structure of the demand model allows us to solve this curse of dimensionality problem and hence apply the insights of the optimal IV literature to the estimation of $\boldsymbol{\Sigma}$. When preferences can be described by a linear function of characteristics and random-coefficients, the reduced-form of the model takes an analytically simplified form that can be identified within a single cross section of products, or a small of panel of markets. The key feature of the characteristics model that we exploit is the existence of fully exchangeable aggregate demand function, for which the product index j itself is not informative once we condition on the product's characteristic $\boldsymbol{x}_{jt}$. We discuss this result in the next section.

2 Structure of Optimal Instruments

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

[5] This problem is relevant in practice, since many important applications use fairly large n_t and small T. For example in the case of the original automobile data (which we will analyze later) we have roughly 100 products in a market with 5 product characteristics, making $\boldsymbol{x}_t$ 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.

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

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

$$\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\Sigma}\right) = f\left(\boldsymbol{\omega}_{jt}; \boldsymbol{\Sigma}\right) + C_t, \quad j = 1, \ldots, n_t \tag{7}$$

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

The proof of this proposition 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: $\tau_j = \exp(\delta_j)/\left(1 + \sum_{j'} \exp(\delta_{j'}\right)$ for all $j = 0, \ldots, n_t$. 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 (7). 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.[6] 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 1. *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

[6] 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 $,\boldsymbol{x}_{jt} -, \boldsymbol{x}_{kt}$ 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 optimal instruments 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 derivative of the quality-assignment function is a symmetric function of the matrix of characteristic differences:*

$$D_{j,k}(\boldsymbol{x}_t) = E\left[\frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\Sigma}^0\right)}{\partial \boldsymbol{\Sigma}_k}\middle|\boldsymbol{x}_t\right] = g_k\left(\boldsymbol{d}_{jt}\right) + c_{t,k}, \quad \forall k = 1, \ldots, dim(\boldsymbol{\Sigma})$$

where $c_{t,k}$ *is a market/parameter 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 6 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 1). This establishes that the expectation of the inverse-demand, $E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\Sigma}^0\right)\middle|\boldsymbol{x}_t\right]$, 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, since equation (6) is simply the derivative of this conditional expectation, the optimal instruments exhibit the same properties.

To understand the usefulness Proposition 2, consider a special case of the model with a single attribute, x_{jt}. In this case, the first order polynomial approximation of the optimal IV can written as follows:

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

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:

$$\begin{aligned} g(\boldsymbol{d}_{jt}) &\approx \sum_{j' \neq j} \sum_{k \neq j} a_{j',k} d_{jt,k} d_{jt,j'} \\ &= a_1 \cdot \left(\sum_{j' \neq j} d_{jt,j'} \right) + a_2 \cdot \left(\sum_{j' \neq j} d_{jt,j'}^2 \right) + a_3 \cdot \left(\sum_{j' \neq j} d_{jt,j'} \right)^2 \end{aligned}$$

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

Note that Propositions 1 and 2 are 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.

2.1 Differentiation IVs

In order to construct relevant instruments, we look for functions of the market-structure menu that can be used to approximate the reduced-form of the model. 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 gradients as in Newey (1990). Since the optimal instruments are symmetric functions, we can use directly the basis functions suggested by the theory of symmetric polynomials to construct relevant instruments. We now illustrate this point using three examples of what we label *Differentiation IVs*.

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):

$$A_j^1(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} d_{jt,j'}^1 \times d_{jt,j'}^l, \ldots, \sum_{j' \neq j} d_{jt,j'}^K \times d_{jt,j'}^l \right\}_{l=1,\ldots,K} \tag{8}$$

where $d_{jt,j'}^k = x_{j't,k} - x_{jt,k}$ measures the difference between product j and j' along dimension k. If the number of characteristics $K = 3$, this formulation includes 9 instrumental variables. 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.

An alternative interpretation of Proposition 2 is that the reduced-form of the model is function of the empirical distribution characteristics differences relative to each product. Since this distribution is observed in the data, it can be summarized using a finite number of moments. This interpretation is especially useful when the number of products of per market is relatively large. For instance, consider discretizing the distribution of characteristic differences along each dimension using L equally spaced grid points. The following basis functions summarize the joint distribution of characteristics differences:

$$A_j^2(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} 1\left(d_{jt,j'}^k < c_1^k\right) \boldsymbol{x}_{j't}, \ldots, \sum_{j' \neq j} 1\left(d_{jt,j'}^k < c_L^k\right) \boldsymbol{x}_{j't}, \right\}_{k=1,\ldots,K} . \tag{9}$$

The element (k, l) in this instrument vector summarizes the characteristics of rivals located to the left of the l^{th} grid point along dimension k. If the number of characteristics $K = 3$ and $L = 5$, this formulation includes 5×9 instrumental variables.

Finally, a special case of the previous formulation consists of 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:

$$A_j^3(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} 1\left(|d_{jt,j'}^1| < \kappa^1\right) \boldsymbol{x}_{j't}, \ldots, \sum_{j' \neq j} 1\left(|d_{jt,j'}^K| < \kappa^K\right) \boldsymbol{x}_{j't}, \right\}, \tag{10}$$

where κ^k is a proximity threshold (e.g. standard-deviation of $x_{jt,k}$ across all markets). When characteristics are discrete, the indicator variables can be replaced by $1(d_{jt,j'}^k = 0)$.

Importantly, this third formulation is robust to variation in the number of products per market. In particular the first two formulations imply that the within market variance in the instruments goes to zero when the number of products per market goes to infinity. This is similar to the weak IV problem pointed out by Armstrong (2015), concerning the so-called "BLP-IV" used to instrument for prices. As long as the product space expands when the number of products per market increases, "local" measures of product differentiation will exhibit variation asymptotically, and will be strong instruments even in large markets.

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 Jacobian function. As we will illustrate in the Monte-Carlo sections below, 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 10. 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.

Finally, a successful approach to construct strong instruments is the optimal IV approximation proposed by Berry et al. (1999). Rather than constructing the conditional expectation in equation (6) via regressions methods, Berry et al. propose to evalute 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. However, it relies on having a consistent initial estimate of the model parameters, which can be problematic when the initial instrument vector is weak. Therefore, understanding what constitutes relevant instruments remains crucial even under this approach.

2.2 Extensions

Demographic variation In many applications, the number and characteristics of products available in each market is fixed, but the distribution of consumer characteristics vary across markets (e.g. Nevo 2001). While this might at first imply that a different of instruments must be use, 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.

Consider the following indirect utility function:

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^{K_2} v_{it,k} x_{j,k}^{(2)} + \varepsilon_{ijt} \tag{11}$$

where $\boldsymbol{v}_{it} = (v_{it,1}, \ldots, v_{it,K_2})$ is distributed in the population according to a market-specific distribution function. If the heterogeneity across markets is such that it is possible to "standardize" the distributions such that $v_{it,k} = \mu_{t,k} + \sigma_{t,k}\nu_{ik}$, where $\boldsymbol{v}_i \sim F_\nu$, then we can write the predicted market shares as *homogenous* functions of market-structure:

$$\sigma_j\left(\tilde{\boldsymbol{X}}_t, \boldsymbol{\delta}_t; F_\nu\right) = \int \frac{\exp\left(\sum_k \nu_{ik}\tilde{x}_{jt,k} + \delta_{jt}\right)}{1 + \sum_{j'=1}^{n_t} \exp\left(\sum_k \nu_{ik}\tilde{x}_{j't,k} + \delta_{j't}\right)} dF_v(\boldsymbol{\nu}_i) \tag{12}$$

where $\tilde{x}_{jt,k} = \sigma_{t,k}x_j$ is the standardized characteristic of product product j, and $\delta_{jt} = \boldsymbol{x}_{jt}\boldsymbol{\beta} + \sum_k \mu_{t,k}x_{j,k} + \xi_{jt}$. Notice that with this transformation, the standardized characteristics vary across markets, and we can construct *Differentiation IVs* as described above.

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

Two sources of variation have been exploited in the literature to construct valid price IVs: (i) ownership structure (e.g. Berry et al. (1995)), and (ii) cost-shifters (e.g. Nevo (2001)). Both cases are valid, since they are independent of the inverse-demand, and correlated with price. We can therefore measure the strength of Differentiation IVs by estimating $\boldsymbol{\gamma}$ via 2SLS instead of OLS. Let $\boldsymbol{w}_{jt}$ denotes a vector of relevant and valid price instruments.

To incorporate these price instruments into our framework, we need to compute exogenous measures of differentiation using a reduced-form pricing equation, rather than observed price levels. This is important since many models used in the literature incorporate a random coefficient on prices to measure for instance income effects.

In this context, our results show that it is important to measure the degree of differentiation along the price dimension in order to predict the inverse-demand function. This can be done in two stages. First, estimate a reduced form pricing equation using observed characteristics, i.e. $\hat{p}_j = E(p_j|x_j, w_j)$. Second, compute L moments of the joint distribution of product characteristic differences by replacing observed prices with $\hat{p}$, i.e. $d^p_{jk} = \hat{p}_j - \hat{p}_k$. This approach is similar to the one proposed by Reynaert and Verboven (2013) to construct their Optimal IV approximation with endogenous prices.

Natural experiments An often expressed criticism of the main identifying assumption, is that firms endogenously choose characteristics (observed *and* unobserved) in response to changes in the structure of the market. This can violate the CMR either because of a selection problem, or because of the existence of a contemporaneous correlation between ξ_{jt} and the attributes of rival products. When such correlation exists, it is typically very difficult to construct valid instruments for every product attributes.

An alternative approach is to look for *natural experiments* that exogenous change the menu

of product characteristics available in markets. 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 see how this kind of exogenous variation can be introduced in the model, consider the following stylized example. Assume that M markets are observed before and after the "exogenous" exit of a product with characteristic vector $\boldsymbol{x}^*$ present in every markets. Let $m \in \{0, 1\}$ indexes the pre and post treatment of an observation, and assume that surviving products have time-invariant characteristics.

The *Differentiation IVs* described in Section 2.1 suggest that the optimal instruments should describe the distribution of characteristic differences relative to all products available in the market, both before and after the exit of product $\boldsymbol{x}^*$. However, if characteristics are endogenous, these instruments are no longer valid. To get around this, we can construct instrumental variables that measure the distance of each product relative to $\boldsymbol{x}^*$, together with rich fixed-effects, to construct the non-linear analog of a Difference-in-Difference (DiD) regression. For instance, let $z_{jm,k} = 1(|x_{jm,k} - x_k^*| < \kappa_k)$ denotes an indicator variable identifying the products that are most likely impacted by the experiment, and $D_t = 1(t = 1)$ denotes an indicator variable equal to one post-experiment. The "first-stage" of the non-linear model can be written as a DiD regression:

$$J_{jmt,k}\left(\boldsymbol{\theta}^0\right) = \mu_{jm} + \tau D_t + \left(\boldsymbol{z}_{jm} \times D_t\right)\boldsymbol{\pi} + u_{jmt}, \tag{13}$$

where $J_{jmt,k}(\boldsymbol{\theta}^0) = \frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)} | \boldsymbol{\Sigma}^0\right)}{\partial \boldsymbol{\Sigma}_k}$ is the "endogenous variable" of the model, μ_{jm} is a product/market fixed-effect, and τ is a post-experiment fixed-effect. The linear parameters $\boldsymbol{\pi}$ measure the impart of the product exit on the slope of the inverse-demand of similar products. This DiD regression can obviously not be run exactly unless we knew the true non-linear parameter. However, since GMM is implicitly linearizing the problem, by using an instrument vector composed of the treatment variable $(\boldsymbol{z}_{jm} \times D_t)$, along with product/market and time fixed effects, the structural parameters are identified solely by the variation generated associated with this DiD regression. Of course, in practice, other continuous functions of the distance to product $\boldsymbol{x}^*$ can be used as excluded instruments to better capture the exogenous variation created by the natural experiment.

3 Detecting Weak Instruments and Practical Considerations

Detecting the presence of weak IVs in non-linear settings is an area of active research in econometrics. In this subsection, we propose two approaches that are easy to implement. Our objective is to propose a "best-practice" for applied work, and discuss the problem of instruments selection.

Our first approach is perform a series of *ex-post* weak IV tests that rely on a linear approximation of the structural model. An important insight from the non-linear IV literature, is that **relevant**

instruments must be highly correlated with the slope of the residual function with respect to the non-linear parameters $\boldsymbol{\Sigma}$. To see this, recall that any non-linear GMM problem can be re-expressed as a sequence of linear IV regressions by taking a first-order approximation of the residual function around the true value of the parameters:

$$\begin{aligned}\xi_{jt}\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) &= \sum_k (\theta_k - \theta_k^0) \frac{\partial \xi_{jt}\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right)}{\partial \theta_k} + \xi_{jt} + \text{Error} \\ &= \boldsymbol{J}_{jt}\left(\boldsymbol{\theta}^0\right)\boldsymbol{b} + e_{jt} \end{aligned} \tag{14}$$

where $b_k = (\theta_k - \theta_k^0)$.

Equation (14) is known as a Gauss-Newton Regression (or GNR). It can be used to construct a Gauss-Newton algorithm to estimate $\boldsymbol{\theta}$. In particular, starting in a neighborhood of $\boldsymbol{\theta}^0$, the algorithm repeatedly estimates (14) to construct the Gauss-Newton updating step (see Davidson and MacKinnon (2001)):

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \left[\boldsymbol{J}(\boldsymbol{\theta}_i)^T \boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}^T \boldsymbol{J}(\boldsymbol{\theta}_i)\right]^{-1} \boldsymbol{J}(\boldsymbol{\theta}_i)^T \boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}^T \boldsymbol{\xi}(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_i + \hat{\boldsymbol{b}}^{\text{iv}}.$$

where $\boldsymbol{W}$ is a weighting matrix, and $\boldsymbol{b}^{\text{iv}}$ is the linear IV estimate of equation (14) using the instrument vector $A_j(\boldsymbol{x}_t)$.[7] The GMM estimate, $\hat{\boldsymbol{\theta}}^{\text{gmm}}$, is implicitly defined locally as the solution to $\hat{\boldsymbol{b}}^{\text{iv}} = 0$. Similarly, the non-linear least-square estimate of $\boldsymbol{\theta}$ can be obtained repeatedly estimating equation 14 by OLS. This leads to biased estimates, because of the correlation between $\partial \sigma_j^{-1}/\partial \Sigma$ and ξ_{jt} (i.e. the Jacobian is function of market shares). We investigate the sign and size of this bias in Appendix A.

Therefore, at the GMM estimate $\hat{\boldsymbol{\theta}}^{gmm}$, the "first-stage" of the structural model corresponds to:

$$J_{jt,k}\left(\hat{\boldsymbol{\theta}}^{gmm}\right) = \boldsymbol{x}_{jt}\boldsymbol{\pi}_k^{(1)} + A_j(\boldsymbol{x}_t)\boldsymbol{\pi}_k^{(2)} + u_{jt,k}, \quad \forall k = 1, ..., |\boldsymbol{\Sigma}|. \tag{15}$$

At the true parameter value, standard weak IV tests available for linear models can be used to test the joint hypothesis that $\boldsymbol{\pi}_k^{(2)} = 0$ for all k, which is valid test of the model being locally identified. Wright (2003) shows that this logic remains valid at the estimated value of the non-linear parameters. In particular, Wright showed that the rank-test proposed by Cragg and Donald (1993) for linear IV problems is a *conservative* test of the null hypothesis of under-identification. For models with one non-linear parameter (i.e. $|\boldsymbol{\Sigma}| = 1$), this test corresponds to the familiar F-test of $\boldsymbol{\pi}^{(2)} = 0$.[8]

Importantly, this discussion suggests that standard weak IV tests available for linear models can

[7] If $W = I$, this algorithm corresponds to the non-linear two-stage least-square estimator proposed by Amemiya (1974).

[8] This test evaluates wether or not the model is locally identified. Since non-linear models can exhibit multiple local solution (especially with weak instruments), it remains essential to ensure that the tests are implemented at the global minimum of the GMM problem.

be used to judge of the strength of instruments in non-linear models. These tests are therefore easy to implement and interpret. For instance, the structural parameters of the model can be estimated using Matlab or other non-linear optimization software. At the solution of the GMM problem, the exogenous and endogenous variables can be exported to a statistical software, and weak IV tests can easily be constructed by estimating the linear GMM problem described in equations (14) and (15).[9]

Our second approach is an *ex-ante* specification test aimed at evaluating the ability of the instruments to reject the *wrong* model. Recall, that the model parameters are weakly identified by $A_j(\boldsymbol{x}_t)$ if the moment conditions are nearly satisfied at the wrong parameters (**?**):

$$E\left[\boldsymbol{A}_j(\boldsymbol{x}_t)\xi_{jt}\left(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\theta}\right)\right]\begin{cases}=0 & \text{If } \boldsymbol{\theta}=\boldsymbol{\theta}^0,\\ \approx 0 & \text{For some } \boldsymbol{\theta}\neq\boldsymbol{\theta}^0.\end{cases} \tag{16}$$

Although this condition is usually not testable, it is feasible in the context of the mixed-logit model to test the validity of a particular "wrong" model: the multinomial logit model. In particular, if $\epsilon_{ij} \overset{iid}{\sim} \text{T1EV}\left(0,1\right)$, the inverse-demand function under the null hypothesis $\boldsymbol{\Sigma}=0$ takes a linear form:

$$\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\Sigma}=0\right)=\ln s_{jt}/s_{0t}=\boldsymbol{x}_{jt}\boldsymbol{\beta}+\xi_{jt}.$$

If the true parameters are such that $\boldsymbol{\Sigma}^0\neq 0$, equation (16) implies that an instrument vector is **weak** if it is unable to reject the null hypothesis of IIA preferences. 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 restriction that the *Differentiation IVs*, which measure the characteristics of rival products, are excluded from the model:

$$\ln s_{jt}/s_{0t}=\boldsymbol{x}_{jt}\boldsymbol{\beta}+A_j(\boldsymbol{x}_t)\boldsymbol{\gamma}+\xi_{jt}. \tag{17}$$

If the CMR is valid, the IIA hypothesis corresponds to the null hypothesis that $\hat{\boldsymbol{\gamma}}=0$. This specification test is easy to implement, and has an important economic interpretation: failing to find a vector $A_j(\boldsymbol{x}_t)$ that rejects the null hypothesis that $\boldsymbol{\gamma}=0$ implies that the data is consistent with the IIA hypothesis (i.e. $\boldsymbol{\Sigma}^0=0$). This test can be constructed using standard linear regressions techniques (i.e. avoiding any non-linear optimization), which makes it very attractive to perform an initial IV selection analysis.

There exists a close connection between the IIA null hypothesis, and the ideal weak IV test (i.e.

[9] For instance the GMM package *ivreg2* in STATA can be used to estimate (14). This involves regressing $\xi_{jt}\left(\boldsymbol{s}_t,\boldsymbol{x}_t;\hat{\boldsymbol{\theta}}^{gmm}\right)$ on $\boldsymbol{J}_{jt}(\hat{\boldsymbol{\theta}}^{gmm})$ using the same instrument vector used to estimate the mode.

projecting the "true" Jacobian vector on instruments). To see this, note that the predicted value of the second term in equation (17) measures the ability of the instruments to explain deviations in the data from IIA preferences:

$$A_j(\boldsymbol{x}_t)\boldsymbol{\gamma} = L(\Delta\xi_{jt}|A_j(\boldsymbol{x}_t))$$

where $L(\cdot)$ is the linear projection operator, and $\Delta\xi_{jt} = \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\Sigma} = 0\right) - \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\Sigma}^0\right)$. If γ is close to zero, the instrument vector is uncorrelated with the *finite difference* between the inverse-demand obtained under logit and the true unobserved quality. In contrast, the instrument vector is weak if it is uncorrelated with the *partial derivative* of the inverse-demand at the true parameter. The strength of the correlation between the instruments and the finite difference (proportional to $|\hat{\boldsymbol{\gamma}}|$) is thus indicative of the strength of the correlation between the instruments and the unknown jacobian function. Therefore the smaller is the p-value associated with the null hypothesis $H_0 : \boldsymbol{\gamma} = 0$, the stronger are going to be the differentiation IVs at identifying $\boldsymbol{\Sigma}^0$.

Of course failing to reject the null hypothesis that $\boldsymbol{\gamma} = 0$ is consistent with two interpretations: (i) the instruments are weak, and (ii) the true model exhibits IIA preferences. Either way, the non-linear model should **not** be estimated with the candidate instrument vector. Conversely, the IIA hypothesis can be rejected if the CMR are not valid, which is not necessarily an indication of weakness.

Therefore, passing the IIA test is a necessary but not sufficient condition to conclude that the instruments are weak (or that the model is different from logit). Because of this, the IIA test should be presented together with a formal specification test of the validity of the instruments. If the model is correctly specified, the instrument vector should be able to detect deviations from the IIA property (i.e. reject $\hat{\boldsymbol{\gamma}} = 0$), **and** fail to reject the over-identification restrictions at the estimated parameter values (i.e "J-test"). If the chosen instruments do not meet one of these two criteria, the results should be interpreted with caution (just like with any IV problem).

4 Monte-Carlo Simulations

In this section we conduct a series of Monte-Carlo simulations to document the weak IV problems associated with commonly used instruments, and demonstrate the ability of *Differentiation IVs* to solve these problems. We start with a simple environment with one dimension of consumer heterogeneity and exogenous characteristics. We use this framework to illustrate graphically the simultaneity problem presents in random-coefficient models, and the consequences of using weak instruments. We then generalize this environment to settings with multiple dimensions of heterogeneity. We focus in particular on the precision and numerical stability improvements associated with using strong instruments. Finally, we introduce a second simultaneity problem into the model (i.e. endogenous prices), and illustrate the construction and performance of *Differentiation IVs* in

this important case.

4.1 Single dimension model

We start with the simplest environment: exogenous characteristics and one random-coefficient. Consider a balanced panel with n products and T markets. Consumers' indirect utility is given by:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt,1} + \beta_2 x_{jt,2} + \sigma_x \eta_i x_{jt,2} + \xi_{jt} + \epsilon_{ijt} = \delta_{jt} + \sigma_x \eta_i x_{jt,2} + \epsilon_{ijt},$$

where $\epsilon_{ij} \overset{iid}{\sim} \text{T1EV}\,(0,1)$ and $\eta_i \sim N(0,1)$. This leads to the following mixed-logit demand function:

$$\begin{aligned} \sigma_{jt}(\boldsymbol{\delta}_t, \boldsymbol{x}_t^{(2)}; \sigma_x) &= \int \frac{\exp(\delta_{jt} + \sigma_x \eta_i x_{jt,2})}{1 + \sum_{j'=1}^{n} \exp(\delta_{j't} + \sigma_x \eta_i x_{j't,2})} \phi(\eta_i) d\eta_i \\ &\approx \sum_{i}^{I} \frac{\exp(\delta_{jt} + \sigma_x \eta_i x_{jt,2})}{1 + \sum_{j'=1}^{n} \exp(\delta_{j't} + \sigma_x \eta_i x_{j't,2})} \omega_i \\ &= s_{jt} \end{aligned}$$

where $\boldsymbol{x}_t^{(2)} = \{x_{1t,2}, \ldots, x_{nt,2}\}$, $\{\eta_i, \omega_i\}_{i=1,\ldots,K}$ are quadrature weights, and s_{jt} is the observed market share of product j.

Weak IV problem

To illustrate graphically the weak IV problem, we simulate a data-set with 20 products and 10 markets. To simplify the illustration, we consider only one characteristic (i.e. $\beta_1 = 0$).[10] From this simulated data-set, we construct two instruments:

$$\begin{aligned} \text{Market IV:} \qquad & A_{jt}^{m}(\boldsymbol{x}_t) = \sum_{j' \neq j} x_{j',t} \\ \text{Differentiation IV:} \qquad & A_{jt}^{d}(\boldsymbol{x}_t) = \sum_{j' \neq j} 1\left(|d_{jt,j'}^{x}| < \text{sd}(x)\right) \end{aligned}$$

We label the first instrument *Market IV* because it summarizes the characteristics of all rivals, irrespectively of their position relative to j. The second instrument, *Differentiation IV*, is a count of the number of competitors located within one standard-deviation of product j. Both instruments are valid since the data has been constructed with exogenous product attributes.

Recall that an instrument is relevant if it is able to reject the "wrong" model. We submit both instruments to this test by regressing the quality assignment obtained under the multinomial logit model (i.e. $\sigma_x = 0$) on an each product's own characteristic and each instrument. This corresponds to the IIA-test discussed earlier. To represent the test graphically we perform this regression in

[10] The other simulation parameters are: $\xi_{jt} \sim N(0,1)$ and $x_{jt} \sim N(0,1)$.

two steps: (i) project $\sigma_{jt}^{-1}(\sigma_x = 0)$ and z_{jt} on an intercept and x_{jt}, (ii) graph both residuals against the instruments.

Figures 1a to 1d illustrate the difference between weak and strong identification. Figures 1a and 1c confirm that the CMR holds in the simulated data: the two instruments are uncorrelated with the residual quality assignment at the *true* parameters (i.e. $\sigma_x = 2$). Figure 1b shows that there also exists no apparent correlation between the sum of rival characteristics and the residual quality assignment evaluated at the $\sigma_x = 0$ (i.e. lns_{jt}/s_{0t}).

In contrast, the number of "close rivals" is negatively correlated with the IIA quality assignment, and explains 11% of the variance in market-shares. The sign of the correlation is very intuitive. Because consumers have heterogeneous tastes for x_{jt}, products with similar characteristics are closer substitutes. Therefore, products that are relatively isolated in the product space have higher market shares on average, than products that face a large number of similar rivals. The multinomial Logit model explains this correlation between shares and market-structure by assigning higher unobserved quality to differentiated products, and low quality to un-differentiated products. This is what explains the negative correlation observed in Figure 1d.

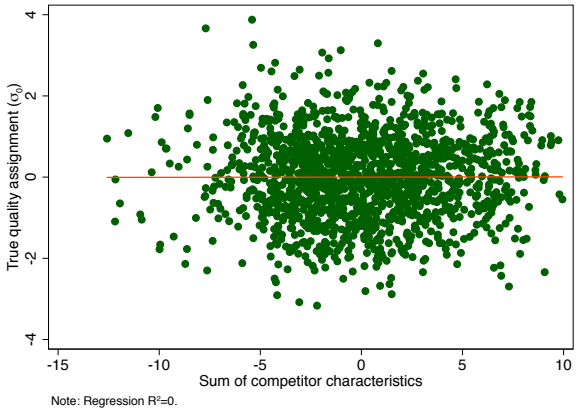
Figure 1b clearly shows that the *Market IV* cannot explain this feature of the market-share data. In particular, since the sum of rival characteristics is uncorrelated the degree of product differentiation, it fails to explain the difference in the distribution of unobserved quality under IIA and non-IIA preferences. Therefore, the *Market IV* is unable to distinguish the model in Figures 1a and 1b, and the random-coefficient parameter is not identified by this moment restriction (or weakly identified).

The two previous examples explain the logic of behind the (lack of) identification of σ_x. To see the consequence of weak identification on the distribution of the GMM estimates, we perform a large-scale Monte-Carlo simulations that repeatedly estimate σ_x using the three estimators.

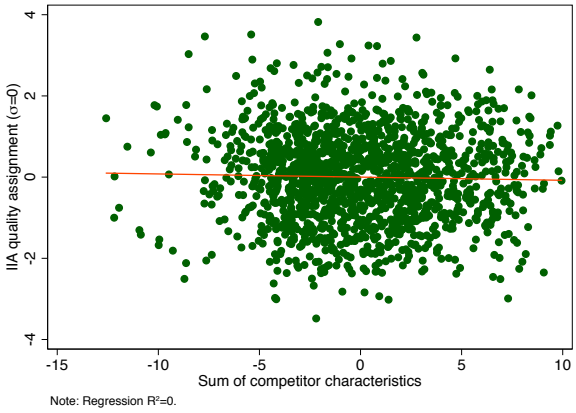
Figure 2 plots the density of three estimates: (i) non-linear least square (long-dash), (ii) GMM with *Market IV* (histogram), and (iii) GMM with *Differentiation IV* (solid). All three densities are constructed using our baseline specification: $n = 100$, $T = 15$, $\text{sd}(\xi) = 1$ and $\beta_0 = -4$. The density of the least-square estimates reflect the simultaneity bias of approximately -40% discussed in Appendix A. In contrast, the density of the GMM estimates obtained with our *Differentiation IVs* is centered around the truth ($\sigma_x^0 = 2$), and has a bell shape consistent with a normal distribution. Notice, that the dispersion of the GMM estimates is slightly larger than the least-square, suggesting that the *Differentiation IVs* produce very precise estimates.

The histogram in 2 confirms that "market-level" instruments produce inconsistent estimates of the random-coefficient parameter. We find that the distribution of $\hat{\theta}^{gmm}$ is highly non-standard, and produces very imprecise estimates. As the figure suggest, nearly 15% of the simulated samples produce estimates equal to zero. When we estimate the log of σ_x, to take into account the corner solution, the average simulated bias is around -200% and the median is -20%; consistent with the

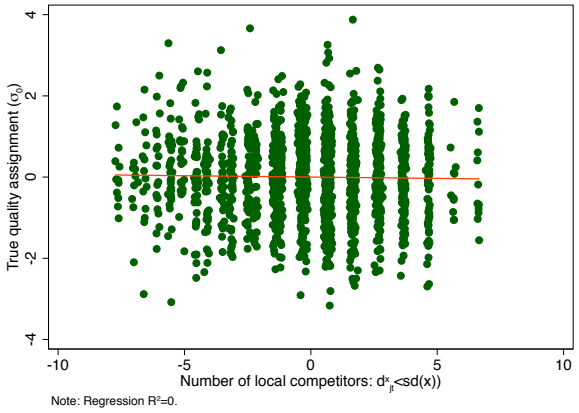
Figure 1: Graphical illustration of the IIA test with exogenous characteristics



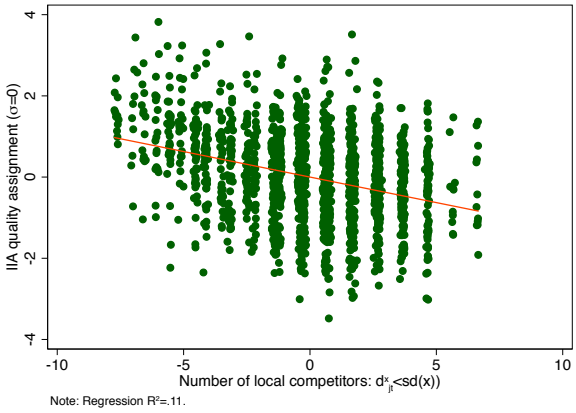
(a) True Quality and Market IV



(b) Logit Quality and Market IV

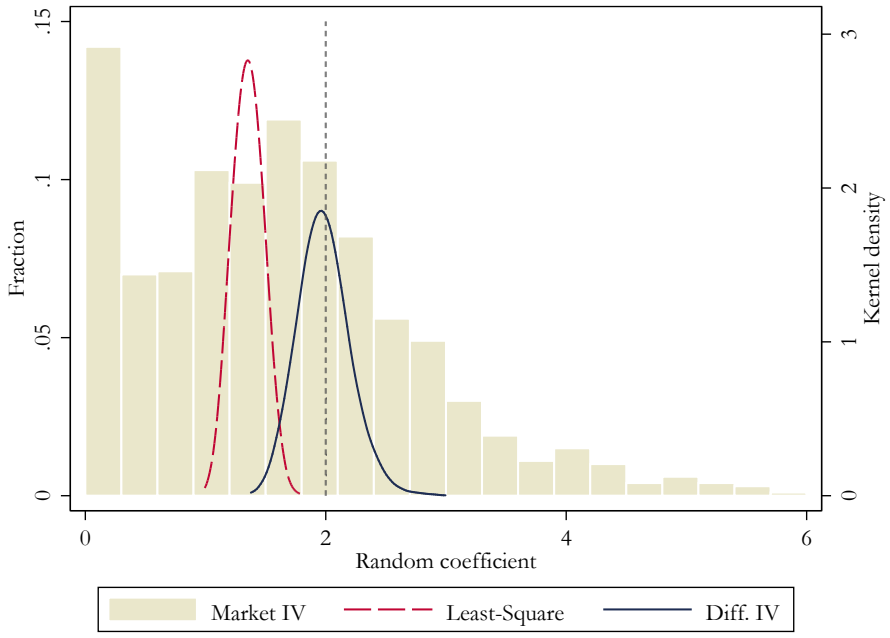


(c) True Quality and Differentiation IV



(d) Logit Quality and Differentiation IV

Figure 2: Simulated distribution of the random-coefficient parameter for three estimating methods



presence of a large number of outliers. In that sense, the NLLS estimates are much more reliable on average than the GMM estimates with weak instruments.

4.2 Multiple dimension model

Next we consider a richer environment with multiple random-coefficients. We use this framework explore the connection between weak identification and difficulties in obtaining reliable and precise estimates. For instance, Reynaert and Verboven (2013) report finding very imprecise results in Monte-Carlo simulations involving four or more random coefficients. Also, Metaxoglou and Knittel (2014) carefully documents a series of numerical problems associated with the estimation of the mixed-logit model (including convergence failure and multiple local minima). In this section, we use the exogenous-characteristic model as a laboratory to reproduce some of these problems, and argue that using "strong" instruments can eliminate most of these problems.

We consider a linear characteristic model with K_2 random-coefficients:

$$u_{ijt} = \beta_0 + \boldsymbol{x}_{jt}\boldsymbol{\beta} + \sum_{k=1}^{K_2} v_{ik} x_{jt,k}^{(2)} + \xi_{jt} + \epsilon_{ijt} = \delta_{jt} + \sum_{k=1}^{K_2} v_{ik} x_{jt,k}^{(2)} + \epsilon_{ijt},$$

where $\epsilon_{ij} \overset{iid}{\sim} \text{T1EV}(0,1)$ and $\boldsymbol{v}_i \sim N(0, \boldsymbol{\Sigma})$. The number of observed product characteristics if given by $K = K_2 + 1$.

This leads to the following mixed-logit demand function:

$$\begin{aligned}
\sigma_{jt}(\boldsymbol{\delta}_t, \boldsymbol{x}_t^{(2)}; \sigma_x) &= \int \frac{\exp\left(\delta_{jt} + \sum_{k=1}^{K_2} v_{ik} x_{jt,k}^{(2)}\right)}{1 + \sum_{j'=1}^{n} \exp\left(\delta_{j't} + \sum_{k=1}^{K_2} v_{ik} x_{j't,k}^{(2)}\right)} f(\boldsymbol{v}_i) d\boldsymbol{v}_i \\
&\approx \sum_{i}^{I} \frac{\exp\left(\delta_{jt} + \sum_{k=1}^{K_2} v_{ik} x_{jt,k}^{(2)}\right)}{1 + \sum_{j'=1}^{n} \exp\left(\delta_{j't} + \sum_{k=1}^{K_2} v_{ik} x_{j't,k}^{(2)}\right)} \omega_i \\
&= s_{jt}
\end{aligned}$$

where $\{\eta_i, \omega_i\}_{i=1,\ldots,K}$ are integration weights, and s_{jt} is the observed market share of product j. We integrate the random-coefficients using a three-points grid along each dimension, located at the three uniform percentiles of the marginal distribution of v_{ik} (i.e. 0.25, 0.5 and 0.75). This allows us to perform relatively quickly large-scale simulations for models with up to six dimensions (i.e. $I = 729$ with 6 dimensions).[11]

We first consider the case of IID random-coefficients. For this example, we contrast two sets of instrumental variables:

$$\begin{aligned}
\text{Market IV:} \qquad & A_{jt}^{m}(\boldsymbol{x}_t) = \left\{ \sum_{j' \neq j} x_{j't,1}, \ldots, \sum_{j' \neq j} x_{j't,K} \right\} \\
\text{Diff. IV:} \qquad & A_{jt}^{d}(\boldsymbol{x}_t) = \left\{ \sum_{j'=0}^{n} \left(d_{jt,j'}^{1}\right)^2, \ldots, \sum_{j'=0}^{n} \left(d_{jt,j'}^{K}\right)^2 \right\}
\end{aligned}$$

where $d_{jt,j'}^{k} = x_{j't,k} - x_{jt,k}$. Both vectors contain K instruments, leading to over-identified models (i.e. df=1). The *Differentiation IVs* are constructed using the sum of square of characteristic differences. Similar results can be obtained by using the number of close-by rivals.

For each model specification, we simulated $1,000$ balanced panels with 15 products and 100 markets. The same parameters were used to simulate all specifications: $x_{jt,k} \sim N(0,1)$, $\xi_{jt} \sim N(0,1)$, $\beta_0 = -5$, $\beta_k = 1$ for all $k = 1, \ldots, K$, and $\sigma_k = 2$ for all $k = 1, \ldots, K_2$. The parameters were estimated using a nested-fixed point algorithm with Simplex and Gauss-Newton optimization routines, and the algorithms were initiated at the true value of the parameters.[12]

[11] We chose to perform our simulation this way rather than using Monte-Carlo integration, because we noticed that having a non-symmetric realized distribution of consumer types tend to create multiple local minimum in the GMM problem. Using symmetric discretization of the distribution solved this problem almost entirely (at least with strong IVs).

[12] The inner-loop convergence criteria for the market-share fixed-point was set to 10^{-14}. To speed up the convergence, we used a Quasi-Newton algorithm with analytical derivatives to solve the fixed-point. All numerical calculations were performed using Ox (Doornik 2001).

Precision and consistency

Tables 1a and 1b summarize the results for the two sets of instruments. The top rows summarize the average parameter estimates (top) and the root-mean-square-error or RMSE (parenthesis).

Overall the results match the ones obtained with the one-dimension model above. The specifications estimated with the *Market IVs* are imprecise and sometimes biased, while the estimates obtained with the *Differentiation IVs* are virtually equal to the truth on average and have small RMSE. Interestingly, the RMSE's do not increase when we add more random-coefficients. Instead, the average bias and precision remains fairly constant. This suggests that the difficulty reported by the prior literature in estimating models with rich unobserved heterogeneity were not necessarily caused by a lack of data, but instead likely caused by weak instruments. Across all specifications, we estimate that the *Differentiation IVs* lead to efficiency gains of an order between 10 and 20.

The bottom two rows summarize the convergence properties of each estimator. The first point to note is that the *Market IVs* specifications were significantly more difficult to estimate than the *Differentiation IVs* ones. Almost systematically, the Gauss-Newton algorithm initiated at the true parameter values failed to converge, and we therefore had to use a Simplex algorithm to perform the estimation. This is a clear indication of weak instruments. An important consequence of weak instruments is that the GMM objective function is poorly approximated by a quadratic function, which cause convergence problems for Newton-type algorithms. Consequently this convergence problem does not occur when we use strong instruments. For all 1,000 simulated samples, the Gauss-Newton algorithm initiated at the true parameter value quickly converged to the minimum (often in less than 5 iterations). Even with six random-coefficients, the optimization algorithm converged in 20 seconds on average.

A related problem with the *Market IV* specification is the difficulty in calculating asymptotic standard errors. The last row of Table 1a indicates that in about 80% of the samples, the rank of the covariance matrix between the instruments and the Jacobian vector is less than the number of parameters (i.e. $\boldsymbol{Z}^T\boldsymbol{J}(\hat{\boldsymbol{\theta}})$), which is another indication of weak instruments.

Optimization problems

To investigate further the link between weak instruments and optimization problems, we repeat the previous Monte-Carlo simulations by initializing the optimization algorithm at 10 randomly selected parameter values. This allows us to investigate the presence of multiple local minima.

The first four rows report the parameter estimates and RMSE, measured using the solution with the lowest objective function. Notice that the statistics for the *Differentiation IV* specifications are identical in Tables 2 and 1b (col. 4), suggesting that the starting values are largely irrelevant when using strong instruments (i.e. most functions exhibit a unique global minimum).

This is not the case with weak instruments. The average *Market IV* estimates reported in Table 2 are more biased less precise than in column (4) of Table 1a. We interpret the results in

Table 1: Monte-Carlo results for the IID model

VARIABLES	Dimensions of consumer heterogeneity			
	1	2	3	4
σ_1	2.012	1.931	1.919	2.055
	(1.3)	(1.423)	(1.383)	(1.467)
σ_2		1.978	1.934	2.055
		(1.302)	(1.313)	(1.412)
σ_3			1.992	2.088
			(1.353)	(1.564)
σ_4				2.106
				(1.494)
Number of consumer types	3	9	27	81
Algorithm	simplex	simplex	simplex	simplex
CPU time (sec)	0.617	4.563	6.891	26.785
Full rank	0.89	0.83	0.78	0.78

(a) Weak instruments: Market IVs

VARIABLES	Dimensions of consumer heterogeneity				
	1	2	3	4	6
σ_1	2.002	1.998	1.994	1.994	1.991
	(.107)	(.109)	(.105)	(.111)	(.115)
σ_2		1.995	1.996	1.992	1.993
		(.108)	(.105)	(.111)	(.115)
σ_3			1.995	1.994	1.991
			(.109)	(.113)	(.116)
σ_4				1.995	1.990
				(.11)	(.115)
σ_5					1.997
					(.118)
σ_6					1.993
					(.118)
Number of consumer types	3	9	27	81	729
Algorithm	newton	newton	newton	newton	newton
CPU time (sec)	0.083	0.117	0.173	0.395	20.130
Full rank	1	1	1	1	1

(b) Strong instruments: Differentiation IVs

Each entry measures the average estimates, and the root-mean-square-error (parenthesis). The true value of each parameter is $\sigma = 2$. Each specification is estimated over 1,000 monte-carlo replications. The distribution of the random coefficients is approximated using a three-points grid in each dimension (i.e. 0.25, 0.5 and 0.75 percentiles). The estimation is initiated at the true value of the parameters. Panel structure: 15 products x 100 markets.

Table 2: Global versus local solutions with 4 IID random-coefficients

	Differentiation IV		Market IV	
	Est.	RMSE	Est.	RMSE
σ_1	1.994	0.111	2.107	1.536
σ_2	1.992	0.111	2.053	1.477
σ_3	1.994	0.113	2.088	1.618
σ_4	1.995	0.110	2.131	1.640
Algorithm	Simplex		Simplex	
CPU time (sec)	0.718		33	
Local optima (fraction)	0.002		0.53	
Global min/Local min	46.6		53	
Coef. variation (local optima)	1.6		11	

The true value of each parameter is $\sigma = 2$. Each specification is estimated over 1,000 monte-carlo replications. The distribution of the random coefficients is approximated using a three-points grid in each dimension (i.e. 0.25, 0.5 and 0.75 percentiles). Local optimal are identified by starting the Simplex algorithm at 10 randomly selected starting values. The global/local ratio and coefficient of variation are calculated in the selected samples exhibiting multiple local optimal. Panel structure: 15 products x 100 markets.

Table 1a as reflecting the "global minimum" solutions, since we initiated the algorithm at the true parameter values. Therefore, identifying the global minimum in the presence of weak instruments would require a very large number of initial starting values (i.e. more than 10), or necessitate the use a global optimization algorithm such as MCMC or Simulated Annealing.

The bottom three rows of Table 2 confirm the importance of multiple local solutions in the weak IV specification. With the *Market IVs*, the Simplex algorithm identified at least two solutions in 53% of the simulated samples. In contrast, we identified multiple solutions in only 2 simulated samples (out of 1,000) with the *Differentiation IVs*. Importantly, these local minima are very different on average. The ratio of the maximum solution over the minimum is equal to 53 on average, and the average coefficient of variation across the estimates is 11. This is consistent with the results documented in Metaxoglou and Knittel (2014).

Finally, even when we use a Simplex algorithm for both IV specifications, the computation time is significantly reduced by the use of stronger instruments. On average, each problem took less than 1 second to converge with *Differentiation IVs*, compared to 33 seconds with *Market IVs*.

Correlated random-coefficients

Next we consider a model with correlated random-coefficients:

$$\boldsymbol{v}_i \sim N(0, \boldsymbol{\Sigma}).$$

Rather than estimating the variance-covariance matrix directly, we estimate the Choleski decomposition: $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{P}'$ where $\boldsymbol{P}$ is a lower diagonal matrix with positive diagonal elements. We simulate a model with 4 random-coefficients, which leads to 10 non-linear parameters. For this experiment, we increase the number of products per market to 50, and leave the number of markets unchanged at 100.

To identify this model, we need to use a richer set of differentiation measures. In particular, it is essential that the instruments capture the correlation in the distribution of characteristics differences across multiple dimensions. In particular we consider all binary interactions between the $K = 5$ dimensions of differentiation:

$$\text{Differentiation IV:} \qquad A_{jt}^{d}(\boldsymbol{x}_t) = \left\{ \sum_{j'=0}^{n} d_{jt,j'}^{1} \times d_{jt,j'}^{k}, \ldots, \sum_{j'=0}^{n} d_{jt,j'}^{K} \times d_{jt,j'}^{k} \right\}_{k=1,\ldots,K}.$$

This leads to 25 instruments, instead of the 5 used in the IID models.

The results are summarized in Table 3. The top panel reports the average estimated covariance coefficients (and RMSE), and the middle panel reports the true covariance matrix used to simulate the model. The simulations confirm the ability of the *Differentiation IVs* to identify rich substitution patterns. The GMM estimates are almost indistinguishable from the true parameters, and the RMSEs are of the same magnitude as the ones reported in the IID models. Notice also that the Gauss-Newton algorithm quickly converges (on average in three seconds per sample). This is very encouraging, since this unrestricted covariance model is estimated on a larger sample (5,000 versus 1,5000 observations) and contains more parameters (10 versus 4) than the IID model presented earlier.

To compare the performance of the *Differentiation IVs* with weak instruments, we construct an estimator based on the optimal IV approximation proposed by Berry et al. (1999). Recall that these instruments are based on the following heuristic:

$$E\left(\frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}^{(2)}; \boldsymbol{\theta}^0\right)}{\partial \theta_k} \middle| \boldsymbol{x}_t \right) \approx \left. \frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}^{(2)}; \hat{\boldsymbol{\theta}}\right)}{\partial \theta_k} \right|_{\xi_{jt}=0}.$$

The challenge with this approach is to obtain an initial estimate $\hat{\boldsymbol{\theta}}$. If we use the estimates obtained with *Differentiation IVs* as starting values, the results would be nearly indistinguishable from the ones in Table 3. Instead, we use random starting values to calculate the analytical Jacobian function at $\xi_{jt} = 0$: $\hat{\theta} \sim N(0, 1)$. Of course this can be viewed as a worst-case scenario, since these parameter values are not even consistent estimates. Our objective however is twofold: (i) we want to investigate the robustness of the heuristic approximation to the starting values, and (ii) we want to compare the results obtained with the *Differentiation IVs* with a potentially weaker set of instruments.

Table 3: Monte-Carlo simulation results for the correlated random-coefficient specification

	$\Sigma_{.1}$	$\Sigma_{.2}$	$\Sigma_{.3}$	$\Sigma_{.4}$
Estimated parameters:				
$\Sigma_{1.}$	4.491			
	(3.136)			
$\Sigma_{2.}$	-2.047	4.394		
	(1.379)	(1.57)		
$\Sigma_{3.}$	2.059	0.747	4.558	
	(.735)	(.41)	(.355)	
$\Sigma_{4.}$	-1.243	-0.389	-0.484	3.089
	(.497)	(.307)	(.264)	(.256)
True parameters:				
$\Sigma_{1.}$	4.064			
$\Sigma_{2.}$	-2.083	4.270		
$\Sigma_{3.}$	1.995	0.689	4.505	
$\Sigma_{4.}$	-1.152	-0.398	-0.462	3.071
Nb. Products	50			
Nb. Markets	100			
CPU time (sec)	191.690			
Failed convergence	0.06			
Singular matrix	0.24			
Algorithm	newton			

Each entry measures the average estimates, and the root-mean-square-error (parenthesis). Each specification is estimated over 1,000 monte-carlo replications. The distribution of the random coefficients is approximated using a three-points grid in each dimension (i.e. 0.25, 0.5 and 0.75 percentiles). The estimation is initiated at the true value of the parameters.

Table 4: Monte-Carlo simulation results for the correlated random-coefficient specification with the optimal IV approximation

	$\Sigma_{.1}$	$\Sigma_{.2}$	$\Sigma_{.3}$	$\Sigma_{.4}$
Estimated parameters:				
$\Sigma_{1.}$	4.491			
	(3.136)			
$\Sigma_{2.}$	-2.047	4.394		
	(1.379)	(1.57)		
$\Sigma_{3.}$	2.059	0.747	4.558	
	(.735)	(.41)	(.355)	
$\Sigma_{4.}$	-1.243	-0.389	-0.484	3.089
	(.497)	(.307)	(.264)	(.256)
True parameters:				
$\Sigma_{1.}$	4.064			
$\Sigma_{2.}$	-2.083	4.270		
$\Sigma_{3.}$	1.995	0.689	4.505	
$\Sigma_{4.}$	-1.152	-0.398	-0.462	3.071
Nb. Products	50			
Nb. Markets	100			
CPU time (sec)	192			
Failed convergence	0.06			
Singular matrix	0.24			
Algorithm	Simplex			

Each entry measures the average estimates, and the root-mean-square-error (parenthesis). Each specification is estimated over 1,000 monte-carlo replications. The distribution of the random coefficients is approximated using a three-points grid in each dimension (i.e. 0.25, 0.5 and 0.75 percentiles). The estimation is initiated at the true value of the parameters. The instruments are calculated by evaluating the analytical jacobian at $\xi_{jt} = 0$ and $\hat{\boldsymbol{\theta}} \sim N(0, 1)$.

The results are presented in Table 4. Although this IV procedure tends to perform slightly better than the *Market IVs* presented earlier, the estimates are much less precise and more biased than the ones obtained with the *Differentiation IVs*. Similarly, the convergence properties of the estimator is much worst. Consistent with the idea using random starting values leads to weak instruments, we frequently encounter convergence problems, and the average computation time per simulated is almost 200 seconds (compared with 3 seconds above).

Therefore we conclude that the performance of this approximation to the optimal instruments is highly dependent on the consistency of the starting values. This conclusion differs slightly form the discussion of those instruments in Reynaert and Verboven (2013). They show, using Monte-Carlo simulations, that the quality of the starting values was largely irrelevant in the context of models with low dimensions and IID random-coefficients. Our results show that this conclusion is no longer valid when number of random-coefficients is large, and/or with correlated unobserved heterogeneity.

4.3 Endogenous prices

In this section, we simulate an equilibrium mixed-logit model with vertical differentiation. Our objective is to illustrate the construction and performance of the *Differentiation IVs* in the presence of a second simultaneity problem. We also compare the performance of the instruments with existing instruments proposed in the literature, including those used in Berry et al. ((1995), (1999)) and Reynaert and Verboven (2013).

We consider the following indirect utility function:

$$u_{ijt} = \begin{cases} \beta_0 + \beta_x x_{jt} + \xi_{jt} - \alpha_i p_{jt} + \varepsilon_{ijt} & \text{If } j > 0 \\ \varepsilon_{i0t} & \text{Else.} \end{cases} \tag{18}$$

where $\alpha_i = \beta_p + \sigma_p y_i^{-1}$ and $\ln(y_i - 1) \sim N(0,1)$ (known), and $\epsilon_{ij} \overset{iid}{\sim} \text{T1EV}(0,1)$. Recall that the parameter values are standardized by the implicit variance of ϵ_{ijt}. Since our objective is to simulate a model that exhibits substitution patterns that are close to a pure-vertical differentiation model (e.g. Bresnahan (1987)), we set the spread parameter to 0.1. This leads to the following parameter values: $\beta_0 = 2.5$, $\beta_x = 10$, $\beta_p = 1$, $\sigma_p = 20$. The distribution of product characteristics are given by: $x_{jt} \sim N(0, 3.5)$, $\xi_{jt} \sim N(0, 1.5)$. We introduce variation across markets in the number of products: $n_t = 40 + e_t$ where $e_t \sim \text{Poisson}(10)$.

We simulate equilibrium prices using a multi-product Bertrand-Nash pricing game (Berry et al. (1995)):

$$\boldsymbol{p}_{jt} = \mathbf{mc}_{jt} + (\Omega_t \cdot \Delta_t)^{-1} \boldsymbol{s}_t, \tag{19}$$

where $\Delta_t = \{\partial s_{jt}/\partial p_{kt}\}$ and $\Omega_{jk,t} = 1$ if product j and k and owned by the same firm. To generate random ownership matrix in each market, we simulate a firm identifier from a Poisson distribution

with mean $n_t/2$. Let $\mathcal{F}_{jt}$ denotes the set of products (other than product j) owned by the same firm as product j.

The marginal-cost function is an increasing function of the quality of products:

$$\ln \text{mc}_{jt} = \lambda_0 + \lambda_1 q_{jt} + \omega_{jt}$$

where $q_{jt} = \beta_0 + \beta_x x_{jt} + \xi_{jt}$ is the quality of product j, and $\omega_{jt} \sim N(0, 0.1)$ an unobserved cost residual.[13] In some econometrics specifications, we will treat ω_{jt} as an observed cost shocks.

As we discussed above, the additional challenge when considering models with endogenous prices, is that the reduced-form of the model is function of the distribution of endogenous *price differences*:

$$E\left(\xi_{jt}|x_{jt}, p_{jt}\right) = E\left(f\left(\left\{s_{kt}, d^p_{jt,k}\right\}\right) + C_t - x_{jt}\boldsymbol{\beta}|\boldsymbol{d}_{jt}\right) \neq 0.$$

To evaluate the reduced-form of the model, in theory one must take into account the joint distribution of prices and shares. This is especially challenging when consumers have heterogeneous price coefficients, since this conditional expectation must incorporate the equilibrium supply relationship.

Since our objective is to identify demand without supply, we rely instead on two alternative linear reduced-form equations relating prices and products' own characteristics:

$$\ln \hat{p}^{\text{cost}}_{jt} = \hat{\pi}^{\text{cost}}_0 + \hat{\pi}^{\text{cost}}_x x_{jt} + \hat{\pi}^{\text{cost}}_\omega \omega_{jt}, \tag{20}$$

$$\ln \hat{p}^{\text{no-cost}}_{jt} = \hat{\pi}^{\text{no-cost}}_0 + \hat{\pi}^{\text{no-cost}}_x x_{jt}, \tag{21}$$

where $\hat{\pi}$'s are OLS regression coefficients. We will use these two conditional expectations alternatively to analyze the performance of the estimators with and without cost shifters.

We use these linear regressions to approximate the reduced-form of the model as follows:

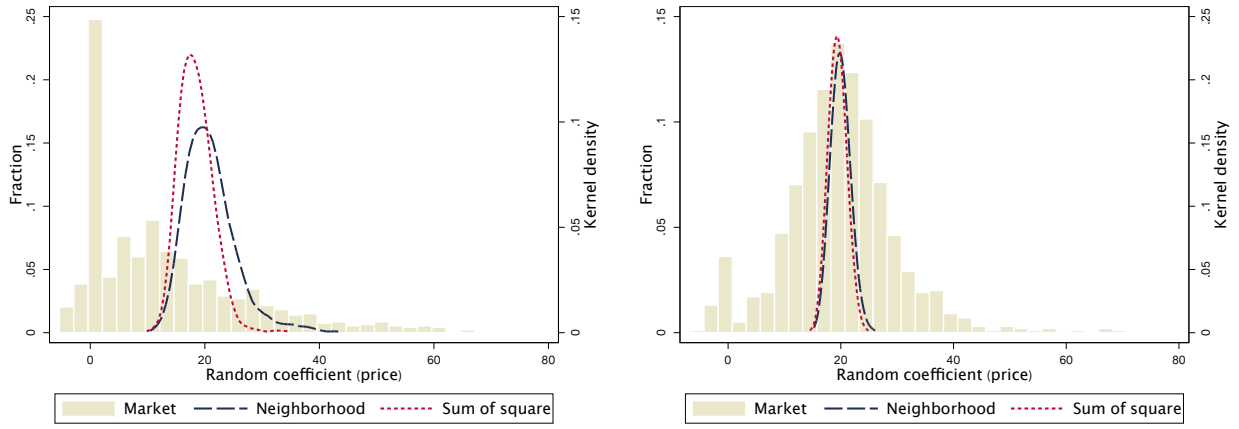
$$D_{j,k}(\boldsymbol{x}_t, \boldsymbol{\omega}_t, \boldsymbol{\Omega}_t) = E\left[\frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{p}_t|\boldsymbol{\Sigma}^0\right)}{\partial \boldsymbol{\Sigma}_k}\middle|\boldsymbol{x}_t, \boldsymbol{\omega}_t, \boldsymbol{\Omega}_t\right] \approx h\left(\left\{d^x_{jt}, d^{\hat{p}}_{jt}, \boldsymbol{\Omega}_t\right\}\right). \tag{22}$$

In other words, rather than explicitly modeling the conditional expectation of the inverse-demand (over prices and shares), we construct our instruments using the conditional expectation evaluated at the expected prices obtained from equation 20 or 21. This is similar to the approach taken by Reynaert and Verboven (2013).

Letting $w_{jt} = \{\text{intercept}, x_{jt}, \omega_{jt}\}$ (i.e. with cost shifter) or $w_{jt} = \{\text{intercept}, x_{jt}\}$ (i.e. without

[13] The parameters of the cost function are: $\lambda_0 = -0.1$, $\lambda_1 = 0.075$.

Figure 3: Distribution of parameter estimates with and without cost shifters



(a) Without cost shifters

(b) With cost shifters

cost shifter), we use three basis functions to construct our instruments:

$$\text{Market IV:} \qquad A_j^m(\boldsymbol{w}_t) = \left\{ \sum_{j' \in \mathcal{F}_{jt}} w_{jt}, \sum_{j' \notin \mathcal{F}_{jt}} w_{jt} \right\}$$

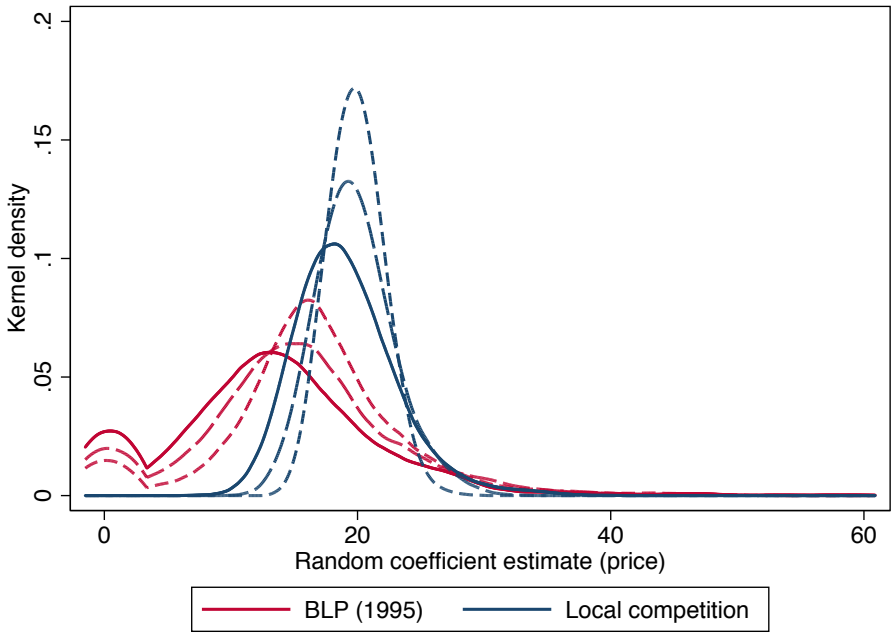
$$\text{Differentiation IV (1):} \qquad A_j^{d1}(\boldsymbol{w}_t) = \left\{ \sum_{j' \in \mathcal{F}_{jt}} \left(d_{jt,j'}^k\right)^2, \sum_{j' \notin \mathcal{F}_{jt}} \left(d_{jt,j'}^k\right)^2 \right\}_{k \in \{x, \hat{p}\}}$$

$$\text{Differentiation IV (2):} \qquad A_j^{d2}(\boldsymbol{w}_t) = \left\{ \sum_{j' \in \mathcal{F}_{jt}} 1\left(|d_{jt,j'}^{\hat{p}}| < \text{sd}(\hat{p})\right) w_{j't}, \sum_{j' \notin \mathcal{F}_{jt}} 1\left(|d_{jt,j'}^{\hat{p}}| < \text{sd}(\hat{p})\right) w_{j't} \right\}$$

The first function corresponds to the instrument vector used in Berry et al. (1995). The second function measures the sum of square of exogenous characteristic differences. The third function measures the sum of exogenous characteristics among rivals located within one standard-deviation in the exogenous price ranking. All three functions measure the characteristics of rivals within and across firms. Overall, we compare the results across six specifications: three measures of differentiation, with and without cost shifters.

Figure 3 plots the distribution of the simulated parameter estimates for the panel with $T = 10$ markets. The results are consistent with the conclusions reached earlier with the exogenous characteristic models. The distributions of the random-coefficient parameter with the *Market IVs* are significantly more dispersed than the distributions obtained with the two *Differentiation IVs*, and share the same non-standard characteristics documented earlier (e.g. bunching at zero, outliers, etc.). Importantly, Figure 3b shows that the inclusion of cost-shifters does not fix this weak identification problem. Although the distribution of the estimated random-coefficient is less dispersed,

Figure 4: Distribution of parameter estimates in small and large samples



Sample size: Solid = 500, Long dash = 1,000, Dash = 2,500.

the *Market IV* estimates are still very dispersed and contain a large number of outliers (e.g. $\sigma_p = 0$ and $\sigma_p > 40$).

Moreover, Figure 4 confirms that the performance of the weak IV specifications does not improve significantly in large samples. In particular, when we increase the number of markets from 10 to 100, the dispersion the parameter estimates with *Market IVs* remains large. To some extent, this is caused by the fact that we observe a large number of outliers even with 2,500 observations (e.g. $\sigma_p = 0$ or $\sigma_p > 40$) . In contrast, the densities of the parameter estimates have a clear bell shape across all three samples, are centered around the true parameter value (i.e. 20).

Table 5a summarizes the simulations results for four specifications: *Market IV* and *Differentation IV*, with and without cost shifters. Consistent with the figures above, we find that the *Market IV* specifications are severally biased and imprecise, especially without using cost shifters. Adding *strong* price instruments (i.e. cost) improves the performance of the *Market IV* specification, but does not eliminate entirely the bias, and the estimates remain very noisy. In contrast, the *Differentiation IV* specifications are precise and unbiased even without using cost shifters, and adding a strong price instrument further improves the precision. Therefore, the results suggest that using a strong price IV is not sufficient to solve the weak identification problems when consumers have heterogeneous price coefficients. It is essential to construct instruments that measure the degree of

Table 5: Monte-Carlo simulations with endogenous prices.

		IV: Sum of charact.		IV: Local competitors	
		w/o cost	w/ cost	w/o cost	w/ cost
β_p					
	Average	0.46	1.08	1.00	1.02
	RMSE	2.19	1.32	0.22	0.18
σ_p					
	Average	13.24	17.47	19.10	19.68
	RMSE	10.84	7.95	3.93	1.51

(a) Market versus Differentiation IVs

		IV: Sum of charact.		IV: Local competitors	
		Market IV	Opt. IV	Diff. IV	Opt. IV
β_p					
	Average	0.46	1.29	1.00	1.16
	RMSE	2.19	0.93	0.22	0.45
σ_p					
	Average	13.24	16.61	19.10	17.28
	RMSE	10.84	28.23	3.93	19.07

(b) Optimal IV approximation without cost shifter

Sample size: $\bar{n}_t \times T = 500$. Number of Monte-Carlo simulations = 1,000. True parameter value: $\sigma_p = 20$, $\beta_p = 1$.

differentiation between products along the exogenous characteristics.

Table 5b investigates the ability of the optimal IV approximation proposed by Berry et al. (1999) to solve this weak identification problem. We consider the specifications without cost shifters, and use the *Market IV* or *Differentiation IV* estimates as starting values to evaluate the Jacobian of the inverse demand with respect to σ_p at $\xi_{jt} = 0$ and $p_{jt} = \hat{p}_{jt}$.

The results obtained using the *Market IV* estimates as starting values show a modest reduction in the bias, but a deterioration in the precision. This latter result is caused by the fact that the optimal IV approximation leads to a outlier parameter estimates. Importantly, the problem of outliers is present both when we use the *Market IV* and *Differentiation IV* estimates as starting values. Removing those samples from the calculation would lead to sizable improvements in the precision of the estimates with the *Market IV* specification, and nearly indistinguishable results with the *Differentiation IV* specifications.

The results from Table 5b confirm that using weak instruments in the first-stage leads to inconsistent and noisy estimates of the non-linear parameter in the second-stage. The problem is exacerbated by the presence of endogenous prices. In particular, our simulations suggest that evaluating the Jacobian function at $\hat{p}$ introduces an additional non-linear approximation error that further deteriorates the precision of the estimates. Improving the precision of the estimates of $\hat{p}_{jt}$, for instance by using a cost shifter, significantly improves the performance of the optimal IV approximation. However, across all of our simulations, we did not find systematic evidence that the approximation improved the performance of the estimator relative to using the *Differentiation IVs* directly.

A Simultaneity bias

To illustrate the need for instruments, we start by documenting the sign and the importance of the simultaneity bias. To do so, we estimate the model by non-linear least-square (NLLS). The expected bias of NLLS corresponds to the expected value of the OLS coefficients of the Gauss-Newton regression:

$$E\left(\hat{\theta}^{nlls} - \theta^0\right) = E\left(\hat{b}^{ols}\right) = E\left[(\boldsymbol{J}^T(\theta^0)\boldsymbol{J}(\theta^0))^{-1}\right] E\left[\boldsymbol{J}^T(\theta^0)\boldsymbol{\xi}\right] \neq 0 \tag{23}$$

The source and strength of this bias depend on the correlation between the slope of the inverse demand and the model residual: $\text{Corr}\left(\partial\sigma_{jt}^{-1}/\partial\sigma_x, \xi_{jt}\right) \neq 0$. As the next simulations will demonstrate, this correlation depends on the degree of substitution between products, and wether or not the market is close or not to a monopolistic competition environment.

Figure 5 plots the distribution of the log difference between the true value of the random-coefficient parameter and the NLS estimate, across different data-generating processes. Each kernel density is constructed using the $1,000$ Monte-Carlo replications.

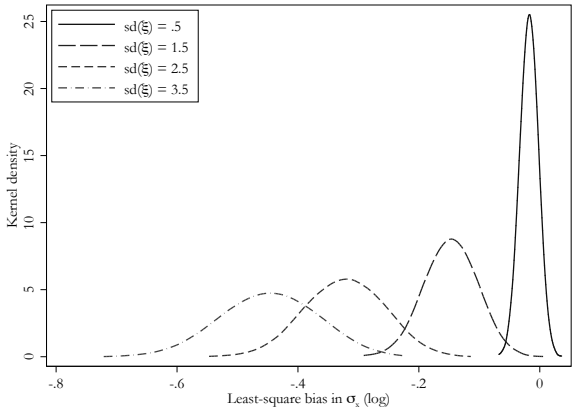
Figure 5a shows that the size of the bias depends intuitively on the variance of the model residual. When the standard-deviation of ξ_{jt} is equal 0.5, the least-square estimate of σ_x essentially unbiased. As we increase $sd(\xi)$, the average bias increases almost linearly, and reaches -45% when $sd(\xi) = 3.5$. As expected, increasing the variance of the model residual also greatly reduces the precision of the estimates.

Interestingly, the sign of the average bias is negative across all of our simulations. In other words, the simultaneity problem systematically leads to an *underestimation* of the amount of the heterogeneity in taste, in favor of the multinomial logit model.

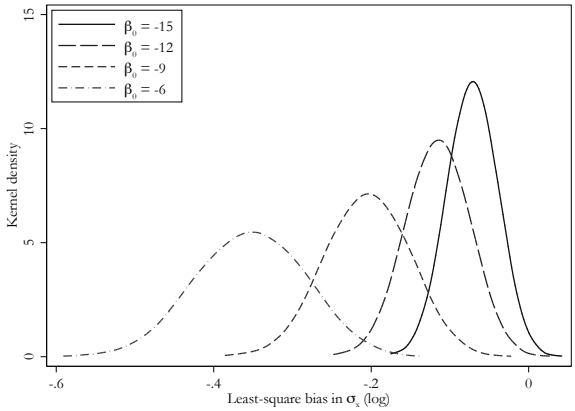
The next two subfigures vary the degree of substitution between products. Figure 5b simulates identical panels with 10 products (on average) and 50 markets, while increasing the value of the outside option; measured by the model intercept β_0. When $\beta_0 = -15$, the outside option is the "closest" substitute of every products, as measured by the diversion-ratio. This is standard feature of mixed-logit models: despite the presence of large heterogeneity in the taste of consumers for observed characteristics, the presence of an idiosyncratic "logit" error implies that elasticity of substitution between products is very sensitive the relative magnitude of the average quality index δ_{jt}. When the quality of products is very small (relative to the outside option), products compete mostly with the outside good, and the derivative of the quality assignment function with respect to σ_x is nearly independent of the characteristics of other products.

To better understand the relationship between the derivative of the quality assignment and the

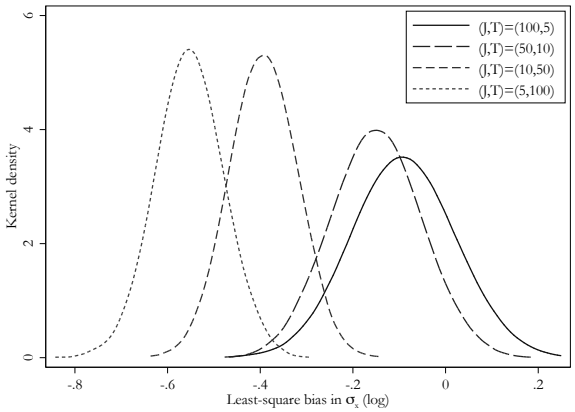
Figure 5: Distribution of simultaneity bias



(a) Variance of the residual



(b) Variance of the model intercept



(c) Panel structure

Each figure plots the Kernel density of the log-difference between the least-square estimate of σ_x and the true parameter value. Each density is estimated using 1,000 Monte-Carlo replications. The default values of the data-generating processes are: $sd(\xi) = 4$, $n = 10$, $T = 50$, $\beta_0 = -10$.

elasticity of substitution, recall that $J_{jt,\sigma_x}(\theta)$ is defined using the implicit function theorem:

$$\frac{\partial \sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \sigma_x\right)}{\partial \sigma_x} = -\left[\frac{\partial \sigma_j\left(\boldsymbol{\delta}_t, \boldsymbol{x}_t^{(2)}; \sigma_x\right)}{\partial \delta}\right]_{j,\cdot}^{-1} \frac{\partial \boldsymbol{\sigma}_t\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \sigma_x\right)}{\sigma_x} \tag{24}$$

Although the last term term varies across products based on σ_x, the $n_t \times n_t$ matrix determining elasticity of demand with respect to the average quality of products contains off-diagonal elements that quickly go to zero when the market share of the outside good increases. When this happens, in our simulation exercise, the variance in $J_{jt,\sigma_x}(\sigma_x)$ across products can be fully explained but the product's own observed non-linear characteristics (i.e. x_{jt}). For instance when $\beta_0 = -15$, H_{jt} is nearly independent of own and competing product market shares, and therefore uncorrelated with the model residual.

As a result, the simultaneity bias from estimating σ_x by least-square is very small when the share of the outside good is large. The four densities confirm that as we increase the average quality of products, the simultaneity bias increases monotonically. When $\beta_0 = -6$, the share of the outside option is about 5%, and the average bias approaches -40%. In contrast, when $\beta_0 = -15$, the share of the outside option is about 90%, and the average bias is less than 5% in absolute value.

Figure 5c illustrates a related implication of the mixed-logit model. When the number of products in each market is small, the degree of "local" substitution is important, and the simultaneity bias is large. This is consistent with the previous discussion. When the number of products grows large, the elasticity of substitution between products become increasingly diffused, and the Jacobian of the quality assignment is nearly independent of products' market shares. In our smallest cross-section example (i.e. $n = 5$), the average bias reaches -55%. In contrast, when the number of products is equal to 100 on average, the the average bias is less than 10% in absolute value.

B Proof of Propositions 1 and 2

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

$$\int \mathbf{1}\left[u\left(t_j, x_j, \theta\right) \geq u\left(t_k, x_k, \theta\right) \ \forall k = 0, \ldots, J, \ k \neq j\right] \ dF\left(\theta\right). \tag{25}$$

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

$$\mathcal{D}_j^{-1}\left(s_0, \ldots, s_J\right) = \mathcal{D}^{-1}\left(s_j, \left\{s_k, d_{jk}\right\}_{k \neq j}\right). \tag{26}$$

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 (25) 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 Doraszleski 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 (26) and hence Theorem 3.

B.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

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. 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. (2015, 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., S. Stern, and M. Trajtenberg (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *The RAND Journal of Economics 28*, s17–s44.

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.

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

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.

Davidson, R. and J. G. MacKinnon (2001). *Artificial Regressions*, Chapter 1. A Companion to Theoretical econometrics. Blackwell, Oxford.

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.

Doornik, J. A. (2001). *Ox - An Object-Oriented Matrix Programming Language*. Timberlake Consultants.

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

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). working paper, UCSD.

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.

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

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

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