Measuring Substitution Patterns in Differentiated Products Industries – The Missing Instruments –

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

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

June 29, 2015

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 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 (2008)). 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 provide a novel empirical strategy for estimating substitution patterns in differentiated product markets that solves these aforementioned difficulties. Our main result shows that there exists an ideal set of instruments in mixed logit demand systems that have been unexploited in empirical work and provide the fundamental source of variation in the data that identifies substitution patterns. We refer to these instruments as "Differentiation IV's" which can be directly constructed in the data and have an intuitive interpretation as measures of the "local market structure" facing a product. We show Differentiation IV's are optimal both consistency and efficiency of measuring substitution patterns in the data. Differentiation IV's enable researchers to clearly and credibly estimate substitution patterns without having to make supply assumptions or appeal to special data sources (such as micro moments, second choice data etc).

We arrive at Differentiation IV's by first isolating the variation in the data that is relevant for identifying structural demand parameters, which builds closely on the recent insights of Berry and Haile (2014) (henceforth BH for short). A folk wisdom that has been "in the air" since BLP is that the correlation between a product's market share and the entry and exit of products with more or less similar characteristics provide the identifying variation for substitution patterns in mixed logit models. We formalize this idea using an intuitive reduced form regression that captures the how the relative market share between two products changes as a function of competing products in the market. We refer to this reduced form object as the IIA regression because it provides a test of the IIA prediction of the simple logit. If the IIA regression exhibits a departure from the simple logit in the data, then the form of the departure is sufficient to non-parametrically identify the parameters governing substitution patterns in the mixed logit using the identification argument in Berry and Haile (2014).

Thus the IIA regression is the fundamental source of variation in the data that identifies consumer heterogeneity and hence substitution patterns. However this also reveals the key challenge for empirical work. The IIA regression requires in principle a very large number of observations on the same product facing exogenously different market structures. This places major burden on the data because market structure is typically a very high dimensional random vector in real applications. Thus the identification strategy in BH can be difficult to apply for market structures with more than just a few products.

Our main result shows that the "characteristics approach" underlying preferences in BLP, which BH do not exploit, gives rise to a sufficient statistic for market structure that preserves the identifying power of the IIA regression yet feasible to apply in the data. We refer to this sufficient statistic as the "local market structure" facing a product, which takes the form of the empirical distribution of differences in product characteristics between a given product and its competitors. Its key advantage for empirical work that it provides a lower dimensional summary of total market structure that is sufficient for identifying the IIA regression (and hence consumer heterogeneity), but does not suffer from a curse of dimensionality as the number of products in a market grows large.

If we take a parametric approach and approximate the IIA regression with flexible class of basis functions, this gives rise to a set of transformations of local market structure that precisely correspond to the relevant instruments for identifying demand in parametric applications. We refer to these transformations as Differentiation IV's. They correspond to intuitive measures of isolation in product space and can be constructed directly in the data but have not yet been systematically used in empirical work. Differentiation IV's are optimal for both constructing consistent and efficient estimators of demand and can be incorporated readily into standard GMM estimation that is already familiar in empirical practice.

We illustrate the usefulness of Differentiation IV's using a series of Monte-Carlo simulations, and by revisiting the car application first studied by Berry et al. (1995). Our results show that the use of Differentiation IVs reduce the small sample bias by a factor of 10, relative to the moment conditions most commonly used in the literature. In addition, we show that similar efficiency gains can be obtained in the context of the U.S. car market, without relying on a particular supply-side assumption, as it is commonly done in the literature.

Related Literature

Our Differentiation 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.[1] However a key point in Berry and Haile

[1] 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.

(2014) is that the identification of substitution patterns poses a distinct empirical problem from price endogeneity.[2] This is because there are in fact *two* different sets of endogenous variables in the model - prices *and* market shares - which require different sources of exogenous variation for the model to be identified. However the literature has been virtually silent about the appropriate form of the instruments for market share? We believe the root of the problems encountered in empirical practice is that there does not exist any formal discussion of how to construct such instruments, and thus researchers have used a single set of instruments, namely price instruments - i.e., instruments constructed on the basis of what should vary price in the model - as instruments for both prices *and* markets shares. This rather naturally has led to a situation where substitution patterns are likely to be at best weakly identified and potentially non-identified.[3] Unfortunately, the parameters governing substitution patterns in the model are non-linear, which makes the issue of constructing strong instruments difficult in general. Our Differentiation IV's can be understood as providing an answer to this issue.

It is important to mention two important exceptions in the literature that have connections to our approach. The first exception are papers that employ a nested logit specification. Nested logit is a special case of the mixed logit model where the problem of endogenous shares is especially clear because shares enter the model takes a linear form, and thus allows for a straightforward application of linear instrumental variable techniques (see e.g., Berry (1994)). Nested logit is however a very strong functional form restriction on substitution patterns, and it is desirable to allow for more flexible patterns of consumer heterogeneity in applications. Unfortunately, the linearity of the nested logit is lost in the general mixed logit model. Our Differentiation IV's can be seen as generalizing the instrumental variable strategy for nested logit models to the larger class of mixed logit models, and in fact collapse back to the standard instruments for market shares in the special case of nested logit (i.e., number of products in the nest). The other exception are papers that estimate spatial demand models where consumers and products are distributed in physical space and the parameter that governs substitution pattens are travel costs (Davis (2006) and Houde (2012)). These papers often employ a measure of the number of local competitors as the instrument to identify travel costs using aggregate market share data. These are types of Differentiation IV's and our main results shows that their effectiveness extends beyond spatial environments to general product characteristic settings.

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

[3] See the Stock and Wright (2000) for a discussion of a related weak identification problem in the Euler equations literature.

1 Identification of the random-coefficient 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) and Berry and Pakes (2002). 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 1.2.

Consider market t with J_t+1 differentiated products. Each product $j = 0, \ldots, J_t$ is characterized by a vector of observed (to the econometrician) product characteristics $\boldsymbol{x}_{jt} = (x_{jt,1}, \ldots, x_{jt,K}) \in \mathbb{R}^K$ and an unobserved characteristic ξ_{jt}. The utility of consumer i for product j is

$$u_{ijt} = \sum_{k=1}^{K} b_{ik} x_{jt,k} + \xi_{jt} + \epsilon_{ijt} \tag{1}$$

where b_{ik} is consumer i's taste for the k^{th} characteristic and ϵ_{ijt} is an idiosyncratic taste for product j. Given the linearity of consumer utility we can normalize the characteristics of an outside good 0 such that $x_{0t} = 0$ and $\xi_{0t} = 0$.[4]

Consumers have heterogeneous tastes in the population and we assume this heterogeneity takes a mixed-logit form. Thus we have that the idiosyncratic taste is distributed $\epsilon_{ij} \overset{iid}{\sim} \text{T1EV}(0,1)$, and the taste for characteristics $\boldsymbol{b}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$, where $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iK}) \overset{iid}{\sim} F_v$ is the heterogenous part of the random coefficient. Let $\left(\boldsymbol{\beta}^0, F_v^0\right)$ denote the true distribution of consumer tastes.

Notice that 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 products across markets. In Section 1.2, 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 $j \in \{0, \ldots, J_t\}$ 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$s_j =$

$$\sigma_j\left(\boldsymbol{X}_t, \boldsymbol{\delta}_t; F_v\right) \quad = \quad \int \frac{\exp\left(\sum_k v_{ik} x_{jt,k} + \delta_{jt}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\sum_k v_{ik} x_{j't,k} + \delta_{j't}\right)} dF_v(\boldsymbol{v}_i) \tag{2}$$

where $\boldsymbol{X}_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{J_t t})$ and $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{J_t t})$. We will refer to $\boldsymbol{X}_t$ as a summary of *market structure* - the number of products and their locations in characteristic space - and will refer to $\delta_t = \boldsymbol{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$ as the average product qualities.

Our empirical problem is to estimate the distribution of consumer taste $(\boldsymbol{\beta}, F_v)$ from an observed

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

panel of shares and product characteristics, i.e.,

$$\text{Data} = \left\{ \{\boldsymbol{s}_{jt}, \boldsymbol{x}_{jt}\}_{j=1}^{J_t} \right\}_{t=1}^{T}, \tag{3}$$

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

An identification problem arises because for any hypothetical distribution $(\boldsymbol{\beta}, F_v)$ there exists a unique unobserved quality vector for products $j = 1, \ldots, J$ that rationalizes the market shares, i.e.,

$$s_{jt} \;=\; \sigma_j(\boldsymbol{X}, \boldsymbol{\delta}_t; F_v) \iff f_{jt}(\boldsymbol{\beta}, F_v) = \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v) - \boldsymbol{x}_{jt}\boldsymbol{\beta}$$

where the inverse σ_j^{-1} uniquely exists following Berry, Gandhi, and Haile (2013). The inverse mapping function f_{jt} corresponds to the structural residual of the model. When evaluated at the true distribution parameters, it is equal to true unobserved quality of products: $f_{jt}(\boldsymbol{\beta}^0, F_v^0) = \xi_{jt}$.[5] Notice that the additive structure of the model implies that the quality assignment function is only function of the distribution of idiosyncratic taste differences, and not of the mean valuations $\boldsymbol{\beta}$.

This indeterminacy implies that $(\boldsymbol{\beta}, F_v)$ cannot be estimated using standard maximum-likelihood techniques. Instead, the econometric problem corresponds to a non-linear instrumental variable regression. Consider the following *structural regression* equation:

$$\ln(s_{jt}/s_{0t}) = \boldsymbol{x}_{jt}\boldsymbol{\beta} + h_j(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v) + \xi_j \tag{4}$$

where $h_j(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v) = \ln(s_{jt}/s_{0t}) - \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v)$ is a *quality assignment function* adjusting the average value of products to reflect heterogeneity in the taste of consumers for product attributes.[6]

This equation makes it clear that a simultaneity problem arises because of the simultaneous determination of market shares. The identification of substitution patterns therefore relies of the ability of the researcher to identify the *causal effect* of competing product market shares, on the products' own relative shares. Since market shares appear on both sides of equation (4), we will refer to this simultaneity problem as a *reflection problem* (Manski 1993). This will allow us to differentiate the standard price simultaneity problem discussed in Berry, Levinsohn, and Pakes (1995), with the endogeneity of market shares limiting our ability to identify F_v.

Equation 4 nests two important cases. When F_v^0 is a degenerate mass distribution at $v_i = 0$, the quality assignment function is zero for all (j, t): $\ln(s_{jt}/s_{0t}) = \boldsymbol{x}_{jt}\boldsymbol{\beta} + \xi_{jt}$. This prediction is driven by the central economic feature of logit preferences - the *independence of irrelevant alternatives*

[5] In order to alleviate the notation burden, we omit the dependence of the residual function on the vector of market shares and product characteristics: $f_{jt}(\boldsymbol{\beta}, F_v) \equiv f_j(\boldsymbol{s}_t, \boldsymbol{X}_t; \boldsymbol{\beta}, F_v)$.

[6] Berry and Haile (2014) consider a different representation of the structural equation. In particular, they consider estimating an *inverse demand* function where the value of each product is expressed in units of a "special regressor". In our context, this would amount to estimate: $\boldsymbol{x}_1 = \frac{1}{\beta_1}\boldsymbol{\sigma}^{-1}(\boldsymbol{s}, \boldsymbol{X}; F_v) - \boldsymbol{X}_{-1}\frac{\boldsymbol{\beta}}{\beta_1} - \frac{1}{\beta_1}\boldsymbol{\xi}_j$, where $\boldsymbol{\sigma}^{-1}(\boldsymbol{s}, \boldsymbol{X}; F_v) = \boldsymbol{\delta}$. Equation 4 is more representative of the specification used in applied work.

(IIA) - which implies that the relative market share between two products is independent of the presence and characteristics of competing products. Because of this relation, we can interpret the quality assignment function, $h_j(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v)$, as measuring the adjustment of product quality away from IIA.

When the random-coefficients interact a series dummy variables representing independent segments, and v_{ik} is distributed according to a type-1 extreme-value distribution, the choice-probabilities take the familiar nested-logit form. In this case, the quality assignment function takes as a closed-form expression:

$$h_j(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v) = \lambda \log(s_{jt|k(j)}),$$

where $s_{jt|k}$ is the observed market share of product j in segment k (Berry 1994).

For more general distributional assumptions, the quality assignment function does not have a closed-form expression, but the reflection problem remains. Thus, identifying the true distribution of consumer heterogeneity from the data requires a restriction on the assignment of qualities ξ_{jt}. We follow Berry, Levinsohn, and Pakes (1995), and employ a restriction with a long history in empirical industrial organization (see e.g., Rosse (1970), Bresnahan (1981, 1987)): we assume that the unobserved quality ξ_{jt} is independent of the market structure $\boldsymbol{X}_t$, i.e.,

$$\text{(CMR)} \quad E[\xi_{jt} \mid \boldsymbol{X}_t] = 0. \tag{5}$$

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. Towards this end we now ask the following central questions:

– What is the variation in the data that identifies the heterogeneity of consumer tastes under the moment restriction (5)?

– How can we use this variation to estimate the distribution both consistently and efficiently from the data (3)?

To answer these two questions, we first discuss the conditions necessary to achieve non-parametric identification of F_v, and then we characterize the construction of valid and relevant instrumental variables required for the estimation of parametric distribution functions.

1.1 Non-parametric identification

When the distribution of the random-coefficients is not specified parametrically, the estimation of equation (4) corresponds to a non-parametric instrumental-variable regression. The structural function $h_j(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v)$ is an unknown function of market shares $\boldsymbol{s}_t$, which are an endogenous variable in the model determined by ξ_{jt}.

The potential instrumental variables are given by the market structure variables $\boldsymbol{X}_t$, as assumed by the CMR (5). In particular, applying the CMR to both sides of equation (4) leads to a *reduced-form* regression function:

$$\begin{aligned} E\left[\ln(s_{jt}/s_{0t}) \mid \boldsymbol{X}_t\right] &= \boldsymbol{x}_{jt}\boldsymbol{\beta}^0 + E\left[h_j\left(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v^0\right) \mid \boldsymbol{X}_t\right] \\ &= \boldsymbol{x}_{jt}\boldsymbol{\beta}^0 + g_j\left(\boldsymbol{X}_t\right). \end{aligned} \tag{6}$$

This reduced-form equation summarizes the variation in the data available to identify F_v. Importantly, if $g_j\left(\boldsymbol{X}_t\right)$ can be identified, then F_v is also non-parametrically identified. In particular, there exists a unique F_v^0 in the model that is consistent with the reduced-form of equation. This can be shown by adopting the same conditions and proof technique as Theorem (1) in Berry and Haile (2014), which we state for completeness and prove in Appendix A. I

This identification result is dependent on our ability to consistently estimate the reduced-form function. Unfortunately, without further restrictions, this task is feasibly **only** when the number of products is small, relative to the number of markets.[7]

The source of the problem is that the non-parametric estimation of equation 6 suffers from a curse of dimensionality: the number polynomial basis terms necessary to approximate $g_j\left(\boldsymbol{X}_t\right)$ grows exponentially with the number of characteristics and the number of products per market. To see this, note that 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. Since identification of $g_j(\boldsymbol{X}_t)$ is a necessary and sufficient condition to identify F_v^0 , this requires the number of products to be small and finite, and the number of markets grows to infinity.

To get around this problem, we now show the the reduced-form of the model as a symmetric function of market-structure. This is not the case for the general preferences studied in Berry and Haile (2014). In their model, there is nothing that ties together different products, and hence each each product j in principle can have its own regression function $g_j(\boldsymbol{X}_t)$.

Fortunately, this problem is not present in the "characteristics model" described in equation (1). When preferences can be described by a linear function of characteristics and random-coefficients, the reduced-form regression (6) takes an analytically simplified form that can be identified within a single cross section of products. The key feature of the characteristics model that we exploit is

[7] This identifying problem is relevant in practice, since many important applications use fairly large J 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.

the existence of an exchangeable aggregate demand function, for which the product index j itself is not informative once we condition on the product's characteristic $\boldsymbol{x}_{jt}$. The next result makes this feature precise.

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. Let us define an ordered pair $\boldsymbol{\omega}_{jt,k} = (s_{kt}, \boldsymbol{d}_{jt,k})$ associated with each product $k = 0, \ldots, J_t$ in the market (including the outside good) for a given inside product $j > 0$, and let $\boldsymbol{\omega}_{jt} = (\boldsymbol{\omega}_{jt,0}, \ldots, \boldsymbol{\omega}_{jt,j-1}, \boldsymbol{\omega}_{jt,j+1}, \ldots, \boldsymbol{\omega}_{jt,J})$. We now have the following result which is proven in the Appendix.

Proposition 1. *Under the linear in characteristics random utility model the quality assignment*

$$h_j \left(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v^0\right) = h \left(\boldsymbol{\omega}_{jt}; F_v^0\right) + c_t, \quad j = 1, \ldots, J_t$$

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

There are two key implications of Proposition 1. The first is that the quality assignment function $h_j \left(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v^0\right)$ 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.[8] The second key implication is that product invariant h is a *symmetric* function of the states of the competing products. Both of these implications give rise to the following key consequence for the reduced-form equation.

Corollary 1. *The function g_{jt} in the reduced-form equation is such that*

$$E \left[h \left(\boldsymbol{\omega}_{jt}; F_v^0\right) \mid \boldsymbol{X}_t\right] = g\left(\boldsymbol{d}_{jt}\right) + c_t$$

where g is a symmetric function of $\boldsymbol{d}_{jt}$ and c_t is a market specific constant.

This corollary shows how we can alleviate the curse of dimensionality in the estimation of the reduced-form. In particular, because the function $g_j\left(\boldsymbol{X}_t\right)$ has reduced to a product invariant function g, this allows us to pool different products together in the data to identify a single function. Most importantly, because g is symmetric in $\boldsymbol{d}_{jt}$ the dimension of the function does not grow unmanageably with the number of products J_t. This implies that the reduced-form of the model is identified under general data-generating processes (DGPs) than the one considered in Berry and Haile (2014). Therefore, this allows to conclude, in Thereom 1, that the the distribution of consumer taste is non-parametric identified. The proof is included in Appendix.

[8] 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.

Theorem 1. *If the CMR holds and preferences can be described by a linear-in-characteristics function (such as in equation 1), the distribution of consumer tastes* $(\boldsymbol{\beta}, F_v)$ *is non-parametrically identified from data on market shares and product attributes when* $\sum_{t=1}^{T} J_t = n \to \infty$.

The symmetry property allows us to take the reduced-form function to the data, and estimate the model in a fashion that is consistent with the non-parametric source of identification. For instance, consider the following two-step minimum distance estimator.

First, estimate the reduced-form function $g(\boldsymbol{d}_j)$, by regressing the log market-share ratios $\log(s_{jt}/s_{0t})$ on a polynomial basis of degree λ describing the joint distribution of product characteristic differences and product j's own characteristics. Let $g_\lambda(\boldsymbol{x}_{jt}, \boldsymbol{d}_{jt}) \approx \boldsymbol{x}_{jt}\boldsymbol{\beta}^0 + g(\boldsymbol{d}_{jt}) + c_t$ denotes this flexible polynomial approximation.

The second step is to estimate a non-parametric distribution of consumer tastes $(\boldsymbol{\beta}, F_v)$ that minimizes the distance between the reduced-form and the predicted quality assignment function:

$$E_\lambda \left[g_\lambda \left(\boldsymbol{d}_{jt} \right) - \boldsymbol{x}_{jt}\boldsymbol{\beta} - h_j \left(\boldsymbol{s}_t, \boldsymbol{X}_t; F_v \right) \middle| \, \boldsymbol{x}_{jt}, \boldsymbol{d}_{jt} \right] = 0, \tag{7}$$

where $E_\lambda[\cdot]$ is the linear projection of the model predictions on a polynomial basis of degree λ.

In practice, it is common to use a parametric distribution of consumer heterogeneity. For instance, a standard approach is to let $v_{ik} = \sigma_k \nu_{ik}$ where $\nu_{ik} \overset{iid}{\sim} N(0,1)$ for $k = 1, \ldots, K$, but more flexible parametric models could be employed given the non-parametric identification of the underlying distribution. The key simplification of the parametric model is that sufficiently rich degree λ polynomial suffices to identify $\boldsymbol{\theta}^0 = (\boldsymbol{\beta}, \boldsymbol{\sigma}^0)$. Since, we do not know ex-ante how large λ will need to be and hence it could be important to be as rich as possible. In particular, we propose to approximate the local market structure facing each product using $N(\lambda)$ moments of the empirical distribution of characteristic differences. Let $\boldsymbol{z}_{jt,1} = \boldsymbol{m}_1(\boldsymbol{d}_{jt}), \ldots \boldsymbol{z}_{jt,N(\lambda)} = \boldsymbol{m}_{N(\lambda)}(\boldsymbol{d}_{jt})$ be the basis functions of this flexible polynomial approximation.

Importantly, the ability of $\boldsymbol{z}_{jt}^\lambda = \left(\boldsymbol{z}_{jt,1}, \ldots, \boldsymbol{z}_{jt,N(\lambda)}\right)$ to identify the parameter vector can be evaluated by measuring its predictive power in the first-stage of the minimum-distance estimation procedure. Corollary 1 makes it clear that if the polynomial basis is not a sufficiently rich description of the local market structure facing each product, it will be uncorrelated with the quality assignment function $h_j\left(\boldsymbol{s}_t, \boldsymbol{X}_t; \boldsymbol{\sigma}^0\right)$. When this happens, the minimum distance estimator is unable to reject the null hypothesis that the distribution of consumer tastes is degenerate at 0, implying that preferences are consistent with the IIA property. As a result, if the true data-generating process has $\boldsymbol{\sigma}^0 \neq 0$, the parameters are only *weakly* identified by $\boldsymbol{z}_{jt}^\lambda$.

Since the polynomial approximation of the reduced-form model takes a linear form, the identifying power of a candidate basis function can easily be evaluated.

Definition 1. *[IIA Test] A vector of characteristics* $\boldsymbol{z}_{jt}$ *weakly identifies the distribution of consumer tastes if it fails to reject the joint hypothesis that* $\boldsymbol{z}_{jt}$ *does not explain the quality assignment* *function. This hypothesis can be tested using a standard specification test:*

$$\begin{aligned} \ln(s_{jt}/s_{0t}) &= \boldsymbol{x}_{jt}\hat{\boldsymbol{\beta}} + \boldsymbol{z}_{jt}\hat{\boldsymbol{\gamma}} + \textit{Residual} \\ W_n(z) &= \hat{\boldsymbol{\gamma}}^T \hat{\boldsymbol{V}}_{\gamma}^{-1} \hat{\boldsymbol{\gamma}}' \end{aligned} \tag{8}$$

where $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ *are obtained via OLS,* $W_n(z) \sim \chi^2(L)$ *is a Wald test measuring the departure of* $\hat{\gamma}$ *from zero,* L *is the dimension of the basis vector* $\boldsymbol{z}_{jt}$*, and* $n = \sum_{t=1}^{T} J_t$ *is the sample size.*

This specification test is easy to implement, and has an important economic interpretation: failing to find a vector $\boldsymbol{z}_{jt}$ that rejects the null hypothesis that $\boldsymbol{\gamma} = 0$ implies that the data is consistent with the IIA hypothesis (i.e. $\boldsymbol{\sigma}^0 = 0$). Moreover, since this test measures the strength of the correlation between $\boldsymbol{z}_{jt}$ and the (unobserved) quality assignment function, the value of $W(z)$ has an analogous interpretation to a "first-stage regression" test commonly used to evaluate the relevance of instruments in linear instrumental variable regression problems. That is, the smaller is the p-value associated with the null hypothesis $H_0 : \boldsymbol{\gamma} = 0$, the stronger are going to be the excluded market-structure variables at identifying $\boldsymbol{\sigma}^0$.

1.2 Differentiation IVs and GMM

It can be shown that the minimum distance estimator described in the previous section is asymptotically equivalent to the GMM estimator proposed by Berry, Levinsohn, and Pakes (1995), where $\boldsymbol{z}_{jt}^{\lambda}$ is used as an excluded instrument vector. We label these transformations of each local market structure as *Differentiation IV's*. Differentiation IV's are statistics of the market structure facing a product, and hence intuitively measure a product's relative isolation in product space.

To see why these variables are *valid* instruments to identify the parameters, observe that the CMR implies that the following unconditional moment restriction is satisfied:

$$E\left[\xi_{jt}\boldsymbol{w}_{jt}\right] = 0 \tag{9}$$

where $\boldsymbol{w}_{jt} = (\boldsymbol{x}_{jt}, \boldsymbol{z}_{jt}^{\lambda})$.

Since the model is non-linear, it is harder to evaluate the *relevance* of the instruments. Chamberlain (1987) shows that the most efficient instrument associated with non-linear parameter σ_k is defined as the conditional expectation of the Jacobian function:

$$z_{jt,k}^{*} = E\left[H_{jt,k}(\boldsymbol{\sigma}^0)|\boldsymbol{X}_t\right]. \tag{10}$$

A simple corollary of Proposition 1 above confirms that this conditional expectation can be written as a symmetric function of the empirical distribution of characteristic differences.

Corollary 2. *The expectation*

$$E\left[\frac{\partial}{\partial \boldsymbol{\sigma}} h_j\left(\boldsymbol{s}, \boldsymbol{X}_t; \boldsymbol{\sigma}^0\right) \mid \boldsymbol{X}_t\right] = H\left(\boldsymbol{d}_{jt}\right) + c_t$$

where H is a symmetric function and c_t is a market specific constant.

This corollary implies that the optimal instruments for $\boldsymbol{\sigma}$ are (unknown) functions of the empirical distribution of characteristic differences relative to each product. In principle, if the true parameters were known, we could therefore approximate this function flexibly using a finite number of moments describing the local market structure around each product (i.e. $\boldsymbol{d}_{jt}$), similar to the minimum distance estimator discussed above.

The corollary also implies that we can use we use these *Differentiation IVs* directly as instruments. This is analogous to the suggestion in Berry et al. (1995) of using the *basis function* as instruments, rather than computing the conditional expectation of the gradients, $E\left[H(\boldsymbol{\theta}^0)|\boldsymbol{d}_{jt}\right]$, using a two-stage approach.

How does this differ from the existing literature? The original instruments proposed by Berry et al. (1995) measure the sum of competing product characteristics, and thus correspond to the first moment of the empirical distribution of characteristic differences. As we will illustrate in the Monte-Carlo section below, these instruments lead to a weak identification of the random-coefficient parameters, because they fail to predict the cross-sectional dispersion in market share. In other words, these variables are essentially uncorrelated with the quality assignment function.

A more successful approach to constructing strong instruments is the optimal IV approximation proposed by Berry et al. (1999). Rather than constructing the conditional expectation in equation (10) via regressions methods, this approach evaluates the Jacobian of the model residual at the unconditional mean of residual (i.e. $\xi_{jt} = 0$), using preliminary estimates of the parameters. 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 instrument vector is weak. For instance, the jacobian of the residual function is zero when evaluated at $\hat{\sigma} = 0$ (i.e. instruments are singular), which is a common realization of GMM with weak instruments as we will in the Monte-Carlo simulations below.

1.3 Extensions and examples

In this section, we illustrate the construction for *Differentiation IVs* for series of common examples found in the empirical literature. We partition the characteristics vector in two groups: $\boldsymbol{x}_{jt}^1$ includes K_1 variables that are not interacted with a random coefficient, and $\boldsymbol{x}_{jt}^2$ is a K_2 dimension vector for which consumers have heterogenous tastes.

Panel data with continuous characteristics For this standard case, we consider two types of moments to approximate the empirical distribution $\boldsymbol{d}_{jt}$. The first one uses a series of discrete *histograms* measuring the distribution of the non-linear characteristics. For each characteristic $x \in \boldsymbol{x}^2$, let $C^x = \{c_1^x, \ldots, c_\lambda^x\}$ denotes λ equally spaced percentiles of the entire distribution of $d_{ij,t}^x = x_{it} - x_{jt}, i \neq j$ pooled across all markets t.[9] For non-linear characteristic x, we construct $\lambda \times K_2$ excluded instrumental variables:

$$\boldsymbol{z}_{jt}^x = \left\{ \sum_{i \neq j}^{J_t} 1(d_{ij,t}^x < c_k^x) \cdot \boldsymbol{x}_{it} \right\}_{k=1,\ldots,\lambda} \tag{11}$$

where $\boldsymbol{x}_{jt} = (\boldsymbol{x}_{jt}^1, \boldsymbol{x}_{jt}^2)$, and $\cdot$ is the element-by-element product indicator. These IVs measure the number and characteristics of competitors located to the left of product j along the x dimension. Note, that these measures of differentiation can also be interacted with product j's own characteristic, to capture the notation that the strength of the correlation between z and the quality assignment varies based on each product's position in the product space.

Depending on the application, it might also be more convenient to characterize the distribution of characteristic differences using continuous moment functions. This is especially relevant when the number of products per market is fairly small, so that the above histograms have little variation. We construct $\lambda \times K_2 \times (K_1 + K_2)$ differentiation measures:

$$\boldsymbol{z}_{jt}^x = \left\{ \sum_{i \neq j}^{J_t} \left(d_{ij,t}^x \cdot (\boldsymbol{x}_{it} - \boldsymbol{x}_{jt}) \right)^k \right\}_{k=1,\ldots,\lambda} \tag{12}$$

Panel data with discrete characteristics In many examples, products are described by a series of discrete product attributes or market segments. This is the case for instance of the nested-logit model discussed in Berry (1994). In this classic example, the quality assignment function takes an analytical form, and the differentiation IVs take the firm of a series of moments characterizing the products available within the same segment:

$$\boldsymbol{z}_{jt}^{\text{NL}} = \left\{ \sum_{i \neq j}^{J_t} \left(1 \left(\boldsymbol{x}_{jt}^2 = \boldsymbol{x}_{it}^2 \right) \cdot (\boldsymbol{x}_{it} - \boldsymbol{x}_{jt}) \right)^k \right\}_{k=1,\ldots,\lambda} \tag{13}$$

where $1(\cdot)$ is an indicator function equal to one if the two products are part of the same nest.

For more general random-coefficient models with discrete attributes, we can similarly construct measures of product differentiation among products sharing the same attributes. This analogous to the histograms defined above, where there is only one distance category: $d_{ij,t} = 1$ if $x_{it} = x_{jt}$.

[9] We construct λ uniform the percentiles between $1/\lambda$ and $(\lambda - 1)/\lambda$.

Large cross-section of products The differentiation IVs defined above exploit the panel variation created by the entry and exit of "similar" products across markets. When the number of products is very large relative to the number of markets, or when only a cross-section of products is available, these variables do not exhibit enough variation.

In this cases, it is essential that the model implies that "local" variation in the characteristics of nearby competitors is more relevant than the overall distribution of characteristic differences, even when the number of products goes to infinity. This condition is satisfied for instance in the nested-logit model, when the number of nests increases in the number of products.

With continuous characteristics, the pattern of differentiation must exhibit strong "local" competition, and the variance of product characteristics must increase when J gets large. This is the case for instance in the quality-ladder model (Bresnahan 1987), or in models of spatial differentiation. Consider for instance the following Hotelling model with logit taste shocks:

$$u_{ij} = \xi_j - \theta(t_i - x_j)^2 + \epsilon_{ij} \quad j = 0, \dots, J$$

where both products $j = 1, \dots, J$ and consumers i have location $x_j, t_i \in [0, 1]$ on a Hoteling line. In this example, like in the general random-coefficient model, the quality assignment function is non-linear.

The main prediction of this model is that the market share of products is increasing in the distance to rivals, everything else being equal. This feature implies that the observed degree of spatial differentiation is a good predictor of the quality assignment away from the true distance cost parameter θ^0. In particular, if we evaluate the residual quality $f_j(\theta)$ at a value of θ that *understates* the disutility of distance for consumers, products that are located in relative crowded space will be assigned *low* quality levels, while products that are isolated will be assigned *high* quality levels. This argument implies that the number of local competitors is a good predictor of the slope of the quality assignment function:

$$z_j^c = \sum_{k=1}^{J} 1\left(\mid d_{jk} \mid < c\right)$$

where c is a distance cutoff. These variables satisfy our definition of Differentiation IV, and vary in the cross-section even when the number of products gets large. These variables correspond to the instruments used in Davis (2006) and Houde (2012) to identify travel costs using aggregate market share data.

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:

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^{K_2} v_{it,k} x_{j,k} + \varepsilon_{ijt} \tag{14}$$

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 $v_{it,k} = \mu_{t,k} + s_{t,k}\nu_{ik}$, where $\boldsymbol{v}_i \sim F_\nu$, then we can write the predicted market shares as *homogenous* functions of market-structure:

$$\sigma_j \left(\tilde{\boldsymbol{X}}_t, \boldsymbol{\delta}_t; F_\nu \right) \quad = \quad \int \frac{\exp\left(\sum_k \nu_{ik}\tilde{x}_{jt,k} + \delta_{jt}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\sum_k \nu_{ik}\tilde{x}_{j't,k} + \delta_{j't}\right)} dF_v(\boldsymbol{\nu}_i) \tag{15}$$

where $\tilde{x}_{jt,k} = s_{t,k}x_j$ 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 product attributes Lastly, Incorporating endogenous prices into the model does not fundamentally change the identification problem of θ, but adds an additional simultaneity problem: in equilibrium prices are correlated with the unobserved quality of products (Berry et al. 1995). We make two observations in this subsection.

Our first observation is that detecting deviations from IIA is not feasible without a separate price instrument. In particular, consider the following linear regression:

$$\ln s_{jt}/s_{0t} = \boldsymbol{x}_{jt}\boldsymbol{\beta} + \alpha p_{jt} + \boldsymbol{z}_{jt}\boldsymbol{\gamma} + e_{jt} \tag{16}$$

where e_{jt} is a composite error term. Since prices are correlated both with the residual and the degree of differentiation (i.e. $\boldsymbol{z}_{jt}$), we can consistently estimate $\boldsymbol{\gamma}$ only using an excluded price instrument.

Two sources of variation have been exploited in the literature: (i) ownership structure (i.e. Berry et al. 1995), and (ii) cost-shifters (i.e. Nevo 2001). Both cases are valid, since they are independent of the quality assignment, and correlated with price. We can therefore measure the strength of Differentiation IVs by estimating $\boldsymbol{\gamma}$ via 2SLS instead of OLS.

Our second observation, is that it is possible 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 quality assignment function. This can be done in two stages. First, estimate a reduced form pricing equation using observed characteristics, i.e. $\hat{p}_j = E(p_j|x_j, w_j)$ where w_j is are cost and/or ownership instruments. Second, compute L moments of the joint distribution of product characteristic differences by replacing observed prices with $\hat{p}$, i.e. $d^p_{jk} = \hat{p}_j - \hat{p}_k$.

2 Monte-Carlo Simulations

In this section we perform a series of Monte-Carlo experiments to illustrate the construction and performance of *Differentiation IVs*. We focus in particular on the consistency and efficiency problems associated with using weak instruments, or with ignoring entirely the simultaneity problem. In the next subsection we quantify the bias associated with simultaneity of competitors' market shares, in the context of a model with exogenous characteristics. Then, in the following subsection, we incorporate endogenous prices into the model, and illustrate the need for two separate sources of exogenous variation.

In all simulations, we consider the following mixed-logit model:

$$u_{ijt} = \underbrace{\beta_0 + \beta_p p_{jt} + \beta_x x_{jt} + \xi_{jt}}_{=X_{jt}\beta + \xi_{jt}} + \sigma_x \eta_i x_{jt} + \varepsilon_{ijt} = \delta_{jt} + \sigma_x \eta_i x_{jt} + \varepsilon_{ijt}, \tag{17}$$

where $\eta_i \sim N(0,1)$ and $\varepsilon_{ij} \overset{\text{iid}}{\sim} \text{T1EV}(0,1)$. The common value of option 0 is normalized to zero (i.e. $\delta_{0t} = 0$). We refer to x_{jt} as the attribute with heterogenous tastes, and p_{jt} as the attribute with homogenous tastes.[10]

These distributional assumptions lead to the following expressions for the market share and quality assignment function of product j:

$$\pi_{jt}(\delta_t, X_t; \sigma_x) = \int \frac{\exp(\delta_{jt} + \sigma_x \eta_i x_{jt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \sigma_x \eta_i x_{kt})} \phi(\eta_i) d\eta_i \tag{18}$$

$$h_{jt}(\sigma_x) = \pi^{-1}_{jt}(s_t, X_t; \sigma_x) \tag{19}$$

$$f_{jt}(\theta) = h_{jt}(\sigma_x) - X_{jt}\beta \tag{20}$$

where $\theta = \{\beta_0, \beta_x, \beta_p, \sigma_x\}$ is the vector of parameter, and δ_t and s_t are $J_t \times 1$ vectors of product quality and observed market shares in market t. The model residual, equation 20, is equal to ξ_{jt} when evaluated at the true parameter vector θ^0.

We generate the data-sets as follows. The number of products per market, J_t, varies across markets according to a poisson distribution: $J_t \sim \text{Poisson}\left(\bar{J}\right)$. Product characteristics are normally distributed: $x_{jt} \sim N(0, v_x^2)$, $p_{jt} \sim N(0, v_p^2)$, $\xi_{jt} \sim N(0, v_\xi^2)$. Except where indicated, the true

[10] The presence of this latter *special regressor* is not crucial to our results and meant here for illustration. This stands in contrast to the crucial role played by the special regressor in Berry and Haile (2014) who don't exploit the same structure of the characteristics model as we do here.

parameter values are: $\beta_0^0 = -5$, $\beta_x^0 = 1$, $\beta_p^0 = 1$, $\sigma_x^0 = 2$. At those parameters, approximately 15% of consumers have a *negative* marginal utility for x_{jt}, and therefore, everything else being equal, would ideally prefer the product with the smallest available x_{jt}. This is a form of market segmentation that is commonly assumed by empirical researchers.

2.1 Exogenous product characteristics

Quantifying the simultaneity bias We start by illustrating the simultaneity bias in σ_x associated with the reflection problem discussed above. To do this, we write the non-linear regression problem using its *artificial regression* form. An artificial regression is a linear regression in which the regressand and regressors are constructed as functions of the data and parameters of a nonlinear model (Davidson and MacKinnon 2001). In our context, this corresponds to a regression of the model residual on its gradients' vector:

$$f_{jt}(\theta) = F_{jt}(\theta)b + \text{Residual} = X_{jt}b_1 + H_{jt}(\sigma_x)b_2 + \text{Residual} \tag{21}$$

where $H_{jt}(\sigma_x) = \frac{\partial h_{jt}(\sigma)}{\partial \sigma_x}$ is the derivative of the quality assignment function.

Equation 21 nests two natural estimators of θ: (1) the non-linear least-square (NLS) estimator is defined as the value of $\theta = \hat{\theta}^{nls}$ such that the OLS estimate of b is equal to zero, and (2) the GMM estimator is defined as the value of $\theta = \hat{\theta}^{gmm}$ such that the IV regression estimate of b is equal to zero. This equivalence comes from the fact that linear regression coefficient vector b corresponds to the score of each non-linear optimization problem, NLS and GMM respectively.[11]

As we discussed in the Identification section, instruments are necessary to correct for the reflection problem. The artificial regression clarifies the existence of this simultaneity problem. Consider estimating the linear regression above by OLS at the true parameter vector (i.e. $\theta = \theta^0$). If the NLS estimator is unbiased, the OLS estimate of b should be zero in expectation. However, since the gradient of the quality assignment is function of the entire vector of market shares, this expectation in general will not be equal to zero:

$$E\left[\hat{b}^{ols} \middle| \theta^0\right] \quad = \quad E\left[(F^T(\theta^0)F(\theta^0))^{-1}\right] E\left[F^T(\theta^0)\xi\right] \neq 0 \tag{22}$$

where $F_{jt}(\theta^0) = \left\{X_{jt}, H_{jt}(\sigma_x^0)\right\}$. The source and strength of this bias depend on the correlation between the slope of the quality assignment and the model residual: $\text{Corr}\left(H_{jt}(\sigma_x^0), \xi_{jt}\right) \neq 0$. Moreover, this correlation depends on wether the quality assignment function is heavily influenced by the market share of a small number of close-by competitors (big bias), or instead mostly a function of products' own attributes (small bias). As we illustrate below, the bias therefore depends

[11] The generalized 2SLS estimate of b is given by: $\hat{b}^{2sls} = \left(F'(\theta)W\Sigma^{-1}W'F(\theta)\right)^{-1} F'(\theta)W\Sigma^{-1}W'f(\theta)$. The second term corresponds to the estimating equation associated with instrument matrix W and weighting matrix Σ, and is therefore equal to zero at $\theta = \hat{\theta}^{gmm}$.

on the degree of substitution between products.

Figure 1 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 1a 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 v_ξ, the average bias increases almost linearly, and reaches -45% when $v_\xi = 3.5$. 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 reflection problem systematically leads to an *underestimate* of the amount of the heterogeneity in taste, in favor of the pure logit model.

The next two subfigures vary the degree of substitution between products. Figure 1b 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.

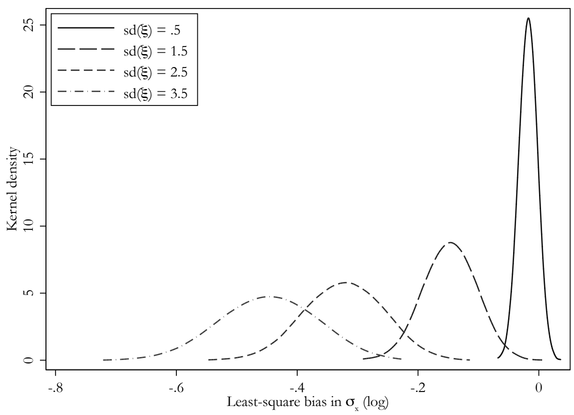
To better understand the relationship between the derivative of the quality assignment and the elasticity of substitution, recall that H_{jt} is defined using the implicit function theorem:

$$H_{jt}(\sigma_x) = \frac{\partial \pi_{jt}^{-1}(\delta_t, X_t; \sigma_x)}{\partial \sigma_x} = -\left[\frac{\partial \pi_t(\delta_t, X_t; \sigma_x)}{\partial \delta_t'}\right]_{j,\cdot}^{-1} \left[\frac{\partial \pi_t(s_t, X_t; \sigma_x)}{\partial \sigma_x}\right]. \tag{23}$$

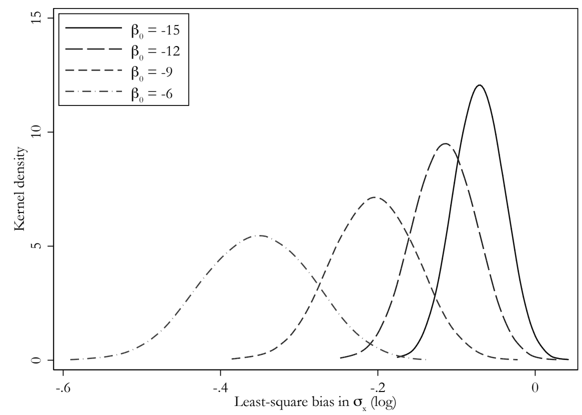
Although the last term term varies across products based on σ_x, the $J_t \times J_t$ matrix determining elasticity of demand with respect to the average quality of products contains off-diagonal elements that quickly go to zero when the market share of the outside good increases. When this happens, in our simulation exercise, the variance in H_{jt} 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 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

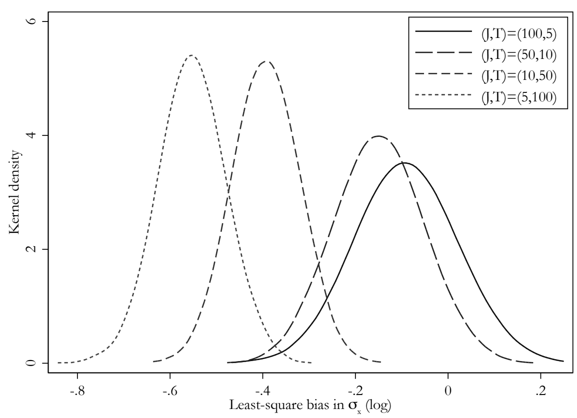
Figure 1: Distribution of simultaneity bias



(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: $v_\xi = 4$, $\bar{J} = 10$, $T = 50$, $\beta_0 = -10$.

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 1c 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. $\bar{J} = 5$), 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.

This last result is similar in spirit to the negative identification results obtained by **?**) in the context of the identification of the price coefficient with weak differentiation IVs.

Eliminating the simultaneity bias The previous section highlights the source of the simultaneity bias in σ_x. To eliminate this bias, it is essential to construct instrumental-variables that are correlated with the slope of the quality assignment, and independent of each product unobserved product quality, ξ_{jt}.

Let $w_{jt} = \{x_{jt}, z_{jt}\}$ denotes a vector of predetermined variables satisfying the conditional independence restriction: $E(\xi_{jt}|w_{jt}) = 0$. It is easy to see that the 2SLS estimate of b is equal to zero in expectation when evaluated at the true value of the parameter vector:

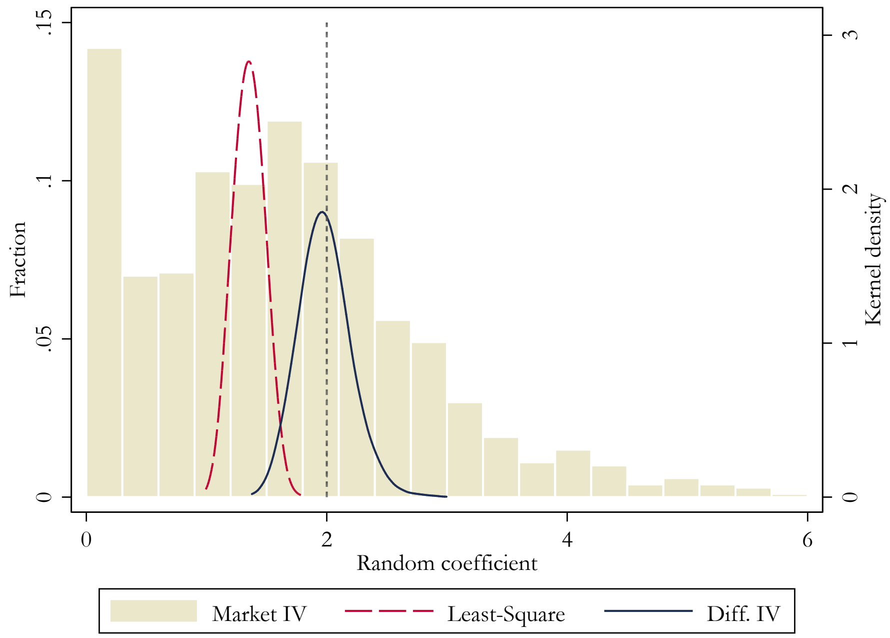
$$E\left[\hat{b}^{2sls}\right] = E\left[\left(F'(\theta^0)W\Sigma^{-1}W'F(\theta^0)\right)^{-1}\right] E\left[F'(\theta^0)W\Sigma^{-1}W'\xi\right] = 0. \tag{24}$$

Since the GMM estimate of θ is defined such that $b^{2sls} = 0$, it is a consistent estimate of the preference parameter vector. However, like any IV estimator, this consistency result crucially depends of the strength of the correlation between w_{jt} and H_{jt}. When w_{jt} is only *weakly* correlated with the slope of the quality assignment function, standard asymptotic theory results break down, and in general $\hat{\theta}^{gmm}$ is not consistent and has a non-standard asymptotic distribution (Staiger and Stock (1997), Stock and Wright (2000)).

Recall, that the best predictor of the *slope* of the quality assignment function is an unknown function of the empirical distribution of characteristic differences between products. Rather than approximating the optimal IV directly, we use a finite number of moments of these differences as instruments. This is analogous to the suggestion in Berry et al. (1995) of using the *basis function* as instruments, rather than directly computing the conditional expectation of the gradients, $E\left[H_{jt}(\theta^0)|w_{jt}\right]$.

We consider two types of moments to approximate $F_j(d)$. The first one uses a series of discrete *histograms* measuring the distribution of the non-linear characteristic, x_{jt}. Let $C^x = \{c_1^x, \ldots, c_K^x\}$ denotes K equally spaced percentiles of the entire distribution of $d_{ij,t}^x = x_{it} - x_{jt}, i \neq j$ across

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



all markets t.[12] For each non-linear characteristic, we construct $K \times |X_{jt}|$ excluded instrumental variables:

$$z^x_{jt,k} = \left\{ x_{jt} \cdot \left(\sum_{i \neq j}^{J_t} 1(d^x_{ij,t} < c^x_k) \cdot X_{it} \right) \right\} \tag{25}$$

where $X_{jt} = \{1, p_{jt}, x_{jt}\}$. These IVs measure the number and characteristics of competitors located to the left of product j along the x dimension. We include the interaction with product j's non-linear characteristic because we found in our simulations that the strength of the correlation varies based on each product position in the product space. In the Monte-Carlo simulations, this approach leads to $K \times 6$ instrumental variables.

In addition, we characterize the distribution of characteristic differences using continuous moment functions. Let l denotes the l^{th} element in vector X_{jt}, and k denotes the power of the function. We construct $K \times |X_{jt}| \times 2$ instrumental variables:

$$z^x_{jt,k} = \left\{ \sum_{i \neq j}^{J_t} x_{jt} (d^x_{ij,t} \times d^l_{ij,t})^k \right\}_{l=1,\dots,|X|} \tag{26}$$

In the Monte-Carlo simulations we use the second and third moments (i.e. $k = 2, 3$). While higher-order moments can yield more precise approximations, they quickly become co-linear when the

[12] We construct K uniform the percentiles between $1/K$ and $(K-1)/K$.

number of products per market is small.

In order to illustrate the consequences of using weak instruments, we construct a third set of instrumental variables that vary only at the market-level. In particular, following Berry, Levinsohn, and Pakes (1995), we use the number of products and the sum of product characteristics within each market:

$$z_{jt}^{\text{Market IV}} = \sum_{j=1}^{J_t} X_{jt}. \tag{27}$$

These variables exploit variation in the number and composition of each simulated markets, but fail to exploit differences across products within each market.

In Table 1 we evaluate the weakness of the differentiation and market IVs. To test the weakness each instrument vector, we construct estimate the "first-stage" of the artificial regression evaluated at the true parameter value:

$$H_{jt}\left(\sigma_x^0\right) = X_{jt}c_1 + Z_{jt}c_2 + e_{jt} \tag{28}$$

where Z_{jt} is an excluded instrument vector (i.e. Market-IVs or Differentiation-IVs).

In practice, this test cannot easily be evaluated since H_{jt} does not have a closed-form expression, and the true parameters are unknown. However, there exists a close relationship between this Weak-IV test, and the IIA-test discussed above. Recall, that we can test the null-hypothesis that $\sigma_x = 0$ estimating the following linear regression:

$$\ln\left(s_{jt}/s_{0t}\right) = x_{jt}\beta + z_{jt}\gamma + \text{Residual} \tag{29}$$

and testing the validity of the exclusion restriction of z_{jt}: $\text{H}_0 : \gamma = 0$. This specification test is easy to implement, and has an important economic interpretation: failing to find a vector z_j that rejects the null hypothesis that $\gamma = 0$ implies that the data is consistent with the IIA hypothesis (i.e. $\theta = 0$).

There exists a close relationship between the two specification tests. In the one-dimension case, the IIA-test is analogous to the first-stage regression in which we replace the partial derivative of the quality assignment, with the discrete change from $\sigma_x = \sigma_x^0$ to $\sigma_x = 0$. In general, this results in a "larger change" in the quality assignment, implying that the IIA null hypothesis is *easier* to reject. The results confirm this intuition. The p-values are on average about twice as large with the first-stage test, as with the IIA-test.

The results also confirm that the Market IVs are very weak instruments. The average p-values in the first column are equal to 0.25 for the first-stage test, and 0.48 for the IIA test. Therefore, in the average simulated sample, we cannot reject the null hypothesis that $\sigma_x = 0$ using the Market-IVs.

The three *Differentiation IVs* are highly correlated with the quality assignment. The first-stage and IIA tests p-values are extremely close to zero in every simulated samples. Both the histogram and the continuous moment specifications appear to explain the quality assignment very well. In

Table 1: Average Weak-IV and IIA specification tests across 1,000 Monte-Carlo simulated sample

	Market IVs	Differentiation IVs		
		Moments	Histogram ($K = 5$)	Histogram ($K = 10$)
IIA-test p-value	0.48	4.42e-13	1.46e-13	5.05e-14
First-stage p-value	0.25	5.33e-18	0	0

Each entry corresponds to the average p-value estimated using 1,000 Monte-Carlo replications. The samples are generated at the default values of the data-generating processes: $v_\xi = 4$, $\bar{J} = 10$, $T = 50$, $\beta_0 = -10$.

the results that we discuss below, we report the GMM estimates obtained with the continuous moments. The results obtained with the histograms yield very similar results, but tend to produce slightly larger biases. We think that this is due to the fact that we use fairly small cross-sections of products, for which the histograms might not produce very accurate approximation of $F_j(d)$.

Figure 2 plots the density of the simulated parameter estimates across three estimators: (i) non-linear least square (long-dash), (ii) GMM with weak instruments (histogram), and (iii) GMM with strong instruments (solid). All three densities are constructed using our baseline specification: $\bar{J} = 10$, $T = 50$, $v_\xi = 3$ and $\beta_0 = -10$. The density of the least-square estimates reflect the simultaneity bias of approximately -40% discussed previously. 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. As predicted by the aforementioned econometrics literature on weak IVs, we find that the limiting 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 presence of a large number of outliers.

Table 2 report the full set of Monte-Carlo results across the same specifications presented in Figure 1. As before each entry is computed over the same 1,000 simulated samples. For each specification and estimator, we report the median bias and the root-mean-square-error (RMSE) of the log difference between the estimated parameter and the true σ_x.[13]

Consistent with Figure 2, the results clearly show that the *Differentiation IVs* successfully eliminate the simultaneity bias in σ_x. The least-square estimator produces median biases that range between -2% and -55%, while the median biases obtained with the strong IVs are all less or

[13]We report the median instead of the mean bias because the average GMM estimate with weak instruments is too much affected by the presence of outliers.

Table 2: Monte-Carlo simulation results with exogenous characteristics

	Least-Square		Market IV		Diff. IV	
	Med. bias	RMSE	Med. bias	RMSE	Med. bias	RMSE
Residual variance (σ_ξ):						
.5	-0.017	0.022	-0.018	0.137	0.000	0.017
1.5	-0.147	0.153	-0.135	4.541	-0.002	0.053
2.5	-0.321	0.328	-0.278	7.252	-0.008	0.088
3.5	-0.448	0.454	-0.332	10.563	-0.018	0.116
Intercept (β_0):						
-15	-0.072	0.077	-0.125	3.820	-0.001	0.037
-12	-0.115	0.121	-0.157	4.826	-0.001	0.047
-9	-0.205	0.211	-0.196	6.034	-0.003	0.064
-6	-0.351	0.359	-0.287	8.055	-0.010	0.093
Panel structure:						
$(\bar{J}=5, T=100)$	-0.552	0.557	-0.207	7.317	-0.031	0.110
$(\bar{J}=10, T=50)$	-0.394	0.399	-0.265	9.015	-0.014	0.101
$(\bar{J}=50, T=10)$	-0.150	0.174	-0.006	7.890	-0.007	0.113
$(\bar{J}=100, T=5)$	-0.094	0.136	-0.004	6.504	-0.009	0.121

Each entry corresponds to median and RMSE of the log-difference between the estimate of σ_x and the true parameter value, over 1,000 Monte-Carlo replications. The default values of the data-generating processes are: $v_\xi = 4$, $\bar{J} = 10$, $T = 50$, $\beta_0 = -10$.

equal to 4% in absolute value (and most less than1%). The RMSE column also confirms that the *Differentiation IVs* produce remarkably precise estimates, even when the variance of the residual is more than three times the variance of product characteristics.

The middle panel summarizes the results obtained with the weak IVs. Consistent with Figure 2, the *Market IVs* produce estimates that are inconsistent and highly dispersed. The median biased are of similar magnitude to the ones found with non-linear least-square. More strikingly, the RMSE of the GMM estimates with weak IVs are about nearly 10 times **larger** than the RMSE of the estimates obtained with the *Differentiation IVs.* This large efficiency gain results from better exploiting the cross-sectional variation in the characteristics of products available in each market.

2.2 Endogenous Prices

We now turn to the more realistic setting in which one of the product attribute is correlated with the unobserved quality of products.[14] In particular, we assume that the linear characteristic x_{2j} measures the price of each product, and is endogenously determined through a Bertrand-Nash pricing game. This model corresponds to a quasi-linear utility function, in which the quality index of product j is defined as:

$$\delta_j = \beta_0 + \beta_1 x_{1j} + \beta_p p_j + \xi_j. \tag{30}$$

[14] Preliminary and incomplete. This section needs to be revised.

If mc_j denotes the constant marginal cost of product j, the price vector is implicitly defined by the following system of first-order conditions:

$$\sigma_j(\delta,\theta) + (p_j - \mathrm{mc}_j)\frac{\partial \sigma_j(\delta,\theta)}{\partial p_j} = 0, \quad \forall j = 1,\dots,J. \tag{31}$$

We will assume that the econometrician observes a cost-shock ω_j that enters the cost function of product j:

$$\ln \mathrm{mc}_j = c_0 + c_1 x_{1j} + \omega_j, \tag{32}$$

where $\omega_j \sim N(0,\sigma_\omega^2)$. In the Monte-Carlo simulations below, we set $c_0^0 = 0$, $c_1^0 = 1$ and $\sigma_\omega = 0.1$.

As discussed in Berry et al. (1995), this equilibrium condition implies that prices are correlated with ξ_j, which tends to bias upward our estimate of β_p if we do not instrument for price or impose additional restrictions. The strength of this correlation, and therefore the magnitude of the bias, depends on multiple factors, including the elasticity of the residual demand of each good, and the variance of the unobserved quality of products. In the Monte-Carlo simulations below, we vary σ_ξ to amplify or attenuate the simultaneity bias (i.e. the larger is σ_ξ, the more serious is the simultaneity bias).

Natural instrument for price is the cost shock ω_j. In order to construct separate instrumental variables to identify θ, as before we construct measures of differentiation along the x_{1j} dimension. In addition, to capture the joint distribution of product attributes, we replace $d_{i,j}^2$ with a measure of cost differential, using the cost shifter of competition products. Using the histogram instrument as an example, our second set of instruments become:

$$\bar{\omega}_j^k = \sum_{j'\neq j} 1\left(d_{j,j'}^1 < c_k\right) \times d_{j,j'}^{\omega}, \quad \forall k = 1,\dots,K, \tag{33}$$

where $d_{j,j'}^{\omega} = \omega_j - \omega_{j'}$ is a cost differential measure between product j and j'. Intuitively, this second instrument measures the cost cost of competitors located in different distance bins of product j.

The presences of an endogenous characteristics changes the GMM estimator as follows:

$$\begin{aligned} \min_{\theta\geq 0} \quad & m_J(\theta) A^{-1} m_J(\theta) \\ \text{s.t.} \quad & m_J(\theta) = \sum_j \left(\delta_j(s,\theta) - x_j\beta(\theta)\right) \times \mathrm{IV}_j \\ & \delta_j(s,\theta) = \sigma_j^{-1}(s,\theta) \\ & \beta(\theta) = \left(x' P_w x\right)^{-1} x^T P_w \delta(s,\theta) \end{aligned} \tag{34}$$

where P_w is the projection matrix using the price instruments. This estimator nests two cases. In the first one, the price instrument vector w_j is a subset of instrument vector IV_j, and includes only the exogenous characteristic and the cost shifter ω_j. In the second case, both instrument vectors are equal (i.e. $\text{IV}_j = w_j$). As we discuss below, the first case is our favorite estimator, and tends to provide better results in our simulations, and in the car application.

Table 3 summarizes the Monte-Carlo simulation results for a cross-section and panel data-generating process (DGP). The cross-section example is similar to the one considered in **??** with 500 products. The panel example also contains 500 observations on average, but products are divided in 25 independent markets. The number of products per market is distributed according to a poisson process with an average of 20 products, and the characteristics of products are independently distributed within and across markets.

Both Tables are divided in three panels, each representing a different combination of moment conditions. In the first two columns, the instrument vector includes the differentiation IVs and the cost shifter, and the linear price coefficients $\beta(\theta)$ are estimated using the cost-shifter as excluded instrument.[15] The second and third columns use only the histogram of characteristic differences as excluded instruments: $\text{IV}_j = w_j = (1, x_{1j}, n_j^1, \ldots, n_j^K)$. Finally, in the third panel, prices are assumed to be exogenous and the instrument vector corresponds to the differentiation instruments: $\text{IV} + j = (1, x_{1j}, n_j^1, \ldots, n_j^K)$ and $w_j = (1, x_{1j}, p_j)$. Within each specification we consider two DGPs: (i) $\sigma_\xi = 1/4$ is associated with a very small correlation between ξ_j and p_j, and (ii) $\sigma_\xi = 1$ is associated with a strong and positive correlation between p_j and ξ_j.

Consider first the simulation results in the cross-sectional example (Table 3a). The first four rows are consistent with the results obtained in the simulations with exogenous characteristics. Across all specifications, the random-coefficient are consistently and precisely estimated. The only specification to yield a somewhat biased estimate of θ is the middle panel in which we do not use a cost-shifter instrument, and the variance of ξ_j is large (i.e. $\sigma_\xi = 1$). This suggests that even when the moment conditions fail to correct for the simultaneity bias in the price coefficient, the differentiation IVs consistently identify the heterogenous taste parameter. In other words, the simultaneity bias appears to be somewhat independent of the reflection problem.

The second section of Table 3a summarizes the results associate with the price coefficient. The main takeaway from this exercise is that having two independence sources of exogenous variation appears to be critical to consistently estimate the price coefficient. The first two columns use the cost shifter to correct for the simultaneity problem, which produces estimates of β_p that are very close to the parameter (i.e. $\beta_p^0 = -2$). This is not the case in the second and third panels, in which the model is estimated solely using the differentiation IVs. With a small σ_ξ, the correlation between p_j and ξ_j is almost zero, and the estimates are close to truth with or without using an instrument for price. With a large σ_ξ, the simultaneity bias is very severe, and the price coefficient is biased upwards with or without using the differentiation IVs to control for simultaneity.

Table 3b shows that the conclusions reached with a cross-section DGP remain valid in the panel. The main difference is that the correlation between p_{jt} and ξ_{jt} is exacerbated with the panel, which

[15] The differentiation IVs include the histogram of characteristic differences n_j^k, and the cost-differential of competitors $\bar{\omega}_j^k$. We set the number of grid points to 10.

Table 3: Monte-Carlo simulation results with endogenous prices

(a) Cross-section

	IVs: Diff. + Cost		IVs: Diff.		No Price IVs	
	$\sigma_\xi = 1/4$	$\sigma_\xi = 1$	$\sigma_\xi = 1/4$	$\sigma_\xi = 1$	$\sigma_\xi = 1/4$	$\sigma_\xi = 1$
Random coef. (θ_x)						
Estimate	0.999	0.997	0.991	0.951	1.001	0.995
Std-error	0.013	0.025	0.054	0.107	0.019	0.077
RMSE	0.019	0.028	0.077	0.134	0.023	0.094
MAE	0.015	0.022	0.062	0.103	0.019	0.075
Price coef. (β_p)						
Estimate	-1.993	-1.975	-1.970	-1.239	-1.965	-1.370
Std-error	0.229	0.234	0.926	0.875	0.209	0.838
RMSE	0.226	0.487	0.910	2.173	0.217	1.057
MAE	0.182	0.374	0.735	1.648	0.171	0.853
J p-value	0.478	0.482	0.478	0.471	0.479	0.478
IIA p-value	0.000	0.000	0.001	0.005	0.000	0.014

True parameters: $\theta = 1$, $\beta_0 = 1$, $\beta_x = 1$, $\beta_p = -2$. Monte-Carlo simulation parameters: nb. simulations = 1,000, $\sigma_{x_1} = 1$, $\sigma_{x_2} = 0$, $\sigma_\xi = 1/4$ or 1, $\sigma_\omega = 0.05$, number of products = 500. Instrumental variables: Histogram of characteristic differences with $K = 10$ grid points (percentiles), Price IV = ω_j. Estimates and standard-errors correspond to the median across all simulated samples.

(b) Panel

	IVs: Diff. + Cost		IVs: Diff.		No Price IVs	
	$\sigma_\xi = 1/4$	$\sigma_\xi = 1$	$\sigma_\xi = 1/4$	$\sigma_\xi = 1$	$\sigma_\xi = 1/4$	$\sigma_\xi = 1$
Random coef. (θ_x)						
Estimate	1.002	1.005	1.000	0.979	0.988	0.812
Std-error	0.019	0.076	0.034	0.137	0.017	0.067
RMSE	0.023	0.093	0.038	0.158	0.022	0.201
MAE	0.018	0.074	0.029	0.120	0.018	0.186
Price coef. (β_p)						
Estimate	-1.991	-1.962	-1.909	-0.572	-1.770	1.447
Std-error	0.222	0.889	0.808	3.267	0.140	0.516
RMSE	0.213	0.853	1.625	6.723	0.297	3.537
MAE	0.168	0.671	1.240	5.122	0.249	3.457
J p-value	0.484	0.484	0.514	0.513	0.453	0.290
IIA p-value	0.000	0.013	0.000	0.014	0.000	0.005

True parameters: $\theta = 1$, $\beta_0 = 1$, $\beta_x = 1$, $\beta_p = -2$. Monte-Carlo simulation parameters: nb. simulations = 1,000, $\sigma_{x_1} = 1$, $\sigma_{x_2} = 0$, $\sigma_\xi = 1/4$ or 1, $\sigma_\omega = 0.05$, number of products = 20, number of markets = 25. Instrumental variables: Histogram of chracteristic differences with $K = 10$ grid points (percentiles), Cost IV = ω_j.

leads to more severe upward biases in the price coefficient. This is because with a smaller choice-set, equilibrium prices exhibit a stronger correlation with the number and characteristics of products. As the number of products increase, the mixed-logit model implies a close to constant markup, which explains most of the differences between Table 3a and 3b.

The presence of this stronger correlation between prices and the unobserved product attribute in the panel case, highlight the importance of instrumenting for prices, for the identification of the random coefficient parameter. The last column shows that treating prices as an exogenous attribute leads to a positive price coefficient, which in turns produce a sizable bias in the estimate of θ. The fourth column shows that this bias in θ is almost entirely eliminated using the differentiation IVs, but the price coefficient is still severely biased (i.e. $-0.572 > -2$).

To better understand the source of the bias in the price coefficient, it is useful to consider the estimation of the linear coefficients $\beta(\theta)$ is GMM problem. Recall, that the empirical moment conditions $m_J(\theta)$ are evaluated by decomposing the quality index using the following 2SLS regression:

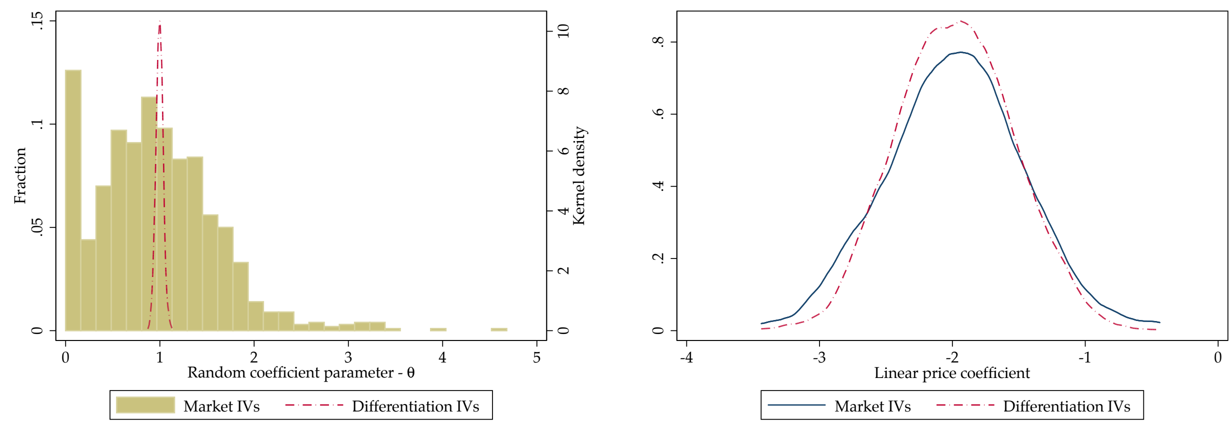
$$\delta_j(s,\theta) = \beta_0^{2sls} + \beta_1^{2sls} x_{1j} + \beta_p^{2sls} p_j + \xi_j^{2sls}(\theta). \tag{35}$$

Recall that for any value of $\theta \neq \theta^0$, the unobserved quality $\xi_j(\theta)$ contains a "specification" error associated the quality assignment and denoted by $\Delta\xi_j(\theta)$. By construction, the differentiation IVs are correlated with $\Delta\xi_j$. Therefore, while the characteristics of similar competing products are relevant instrument for price, they do not satisfy the exclusion restriction, unless the quality index is evaluated at the true parameter. This discussion suggests that any bias in θ can invalidate the use of differentiation IVs as price instruments, and therefore highlight the importance of exploiting two independent sources of exogenous variation to tackle the reflection and simultaneity problems. Our simulation results suggest this bias can arise even from a well specified econometrics model.

Our interpretation of the results in Table 3a and 3b is that the small-sample bias in θ invalidates the exclusion restriction, and creates a systematic upward bias in β_p. In principle this bias should therefore be eliminated asymptotically as the number of products per market or as the number of markets goes to infinity. The results presented in this section exploit fairly large sample, suggesting that the rate at which this bias is eliminated can be very slow.

Finally, Figures 3a and 3b report the distribution of the two key parameter estimates using weak and strong differentiation IVs. In particular, we construct a weak instrument using the number of products and the sum of competing product characteristics. These instruments are labelled "Market IVs" in Figure 3, and correspond to the same instrument proposed by Berry et al. (1995).

In the panel-data case, the Market IVs exploit variation in the number and characteristics of products across markets. Importantly, these variables do not exploit variation in the degree of differentiation between products, and are therefore weakly correlated with the quality assignment function. Similarly to the random product attributes used to construct the weak instrument in Figure **??**, we fail to reject the null hypothesis of IIA with the Market IVs.



(a) Distribution of $\hat{\theta}$

(b) Distribution of $\hat{\beta}_p$

Figure 3: Distribution of parameter estimates with weak and strong IVs in the panel case

Figure 3a confirms that failing to reject the null hypothesis of IIA leads to very imprecise estimates of θ. The distribution of $\hat{\theta}$ is very dispersed, and exhibit the same departures from normality documented in the previous section: (i) nearly 15% of estimates are equal to 0, and (ii) the distribution is highly skewed to the right. The dashed line clearly shows that these two features are not present when the model is estimated using the histogram of characteristics differences.

Finally, Figure 3b shows that the price coefficient is largely unaffected by the choice of differentiation IV, as long as the instrument vector contains a cost-shifter.

3 Empirical Application: Demand for Cars

In this section, we revisit the application of Berry et al. (1995), and measure the degree of differentiation between new cars. Our objective is to illustrate the ability of our differentiation IVs in precisely estimating the parameters of a mixed-logit model without relying on a particular equilibrium model.

Table 4 summarizes the data. The first column shows that the number of models has been steadily increasing between 1971 and 1990, from 92 to a peak of 150 in 1988. The sample size is therefore larger than the ones used in our Monte-Carlo simulation. However, the presence of multiple dimensions of heterogeneity makes the estimation more challenging.

Columns (3) to (8) report the sales weighted average of the main characteristics. This period exhibit important changes in the characteristic and price of new cars. In the particular, while the first oil shock led to a rapid increase in the cost per mile, the second oil shock led to a complete reshuffling of new cars available in the US. This is characterized by a rapid increase in the market share and number of models offered by Asian manufacturers, and a steep decline in the size of cars purchased. The steady decline in oil prices and economic growth in the 1980s contributed

Table 4: Summary statistics of the car data

Year	Models	Price × 1,000	Euro.	Asia	HP/WT / 100	Size × 10,000	\$/Miles /10
1971	92	7.87	0.08	0.06	4.90	1.50	5.60
1972	89	7.98	0.07	0.04	3.91	1.51	5.63
1973	86	7.53	0.03	0.04	3.64	1.53	5.77
1974	72	7.51	0.06	0.05	3.47	1.51	7.26
1975	93	7.82	0.06	0.08	3.37	1.48	7.06
1976	99	7.79	0.04	0.08	3.38	1.51	6.55
1977	95	7.65	0.05	0.11	3.40	1.47	6.03
1978	95	7.64	0.04	0.11	3.46	1.40	5.52
1979	102	7.60	0.04	0.16	3.48	1.34	6.36
1980	103	7.72	0.04	0.19	3.50	1.30	7.16
1981	116	8.35	0.05	0.21	3.49	1.29	6.77
1982	110	8.83	0.05	0.23	3.47	1.28	5.77
1983	115	8.82	0.05	0.21	3.51	1.28	5.08
1984	113	8.87	0.04	0.18	3.61	1.29	4.95
1985	136	8.94	0.05	0.19	3.72	1.26	5.15
1986	130	9.38	0.05	0.22	3.79	1.25	3.67
1987	143	9.97	0.05	0.25	3.95	1.25	3.72
1988	150	10.07	0.05	0.24	3.96	1.25	3.55
1989	147	10.32	0.05	0.26	4.06	1.26	3.71
1990	131	10.34	0.04	0.28	4.19	1.27	4.09

to a reversal of these trends, characterized by increasing real prices, lower fuel efficiency, and the popularity of more powerful engines.

The original model considered by Berry et al. (1995) is a mixed-logit random utility model with standard-normal random coefficients.[16] To illustrate the performance of our Differentation IVs relative to other IVs considered in the literature, we consider the following relatively parsimonious specification of the indirect utility of consumers:

$$u_{ijt} = x_{jt}\beta_x + \mu_j + \tau_t + p_{jt} \cdot (\beta_p + \sigma_p y_{it}) + \sum_{k=1}^{K} \sigma_k \eta_{ik} x_{jt}^k + \xi_{jt} + \varepsilon_{ijt}, \tag{36}$$

where $x_{jt}^k \in \{\text{HP/WT}, \text{DFI}, \text{Four cyl}, \text{AT}\}$, $y_{it} \sim N(\bar{y}_t, \sigma_t^2)$ denotes the log annual income of household i, and (μ_j, τ_t) measures company and period fixed-effects. In addition to the price variable, we allow four variables to exhibit heterogenous tastes: hp/wt, diesel engine, four-cylinder, and automatic transmission.

We consider the same continuous product attributes to describe the willingness to pay of con-

[16]The distribution of the random taste parameter associated with the price variable corresponds to a log-normal distribution estimated using the year-specific income distribution.

sumers: list price ($1983/1,000), car size (i.e. width $\times$ length /10,000), horsepower to weight ratio, and the number of miles per dollar. We augment the model with additional characteristics: air-conditioning, automatic transmission, power-steering, front-wheel drive, four-cylinder, diesel engine. In addition, we estimate a log-linear indirect utility function, rather than expressing product characteristics in levels.

3.1 Identifiation and Instruments Choice

The main identifying assumption is that the unobserved quality index of products is conditionally independent of the characteristics of products available in the market:

$$E(\xi_{jt}|\tau_t, \mu_j, x_{1t}, \ldots, x_{J_t,t}) = 0 \tag{37}$$

where τ_t and μ_j are year and company fixed-effects respectively, and x_{jt} is a vector of pre-determined characteristics of product j.

Our estimation strategy is to construct two sets of instrumental variables consistent with this conditional moment restriction, in order to identify the non-linear parameters $\theta = \{\sigma_p, \sigma_1, \ldots, \sigma_K, \}$ determining the degree of differentiation between products, and the linear price coefficient β_p allowing us to decompose the average valuation of consumers into a quality and a price effects.

As discussed above, the challenge is to exploit exogenous variation in the characteristics of products to solve a reflection and a simultaneity problem. We describe in turn the three class of instrumental variables that we use in the empirical analysis: (i) within firm sum of product characteristics, (ii) product-level cost shifter, and (iii) differentiation IVs.

Berry et al. (1995) consider two types of excluded instrumental variables to get around both problems:

$$\bar{x}_{jt}^{\text{own}} = \sum_{k \in \mathcal{F}_{kt}} x_{kt}, \quad \bar{x}_{jt}^{\text{other}} = \sum_{k \notin \mathcal{F}_{kt}} x_{kt}$$

where $\mathcal{F}_{jt}$ denotes the set of competing products sold by the same manufacturer as product j in period t. These two variables are defined over the main discrete and continuous characteristics: hp/wt, size, air-conditioning, and the intercept. Since our econometric specifications include year fixed-effects, we cannot use both variables (i.e. $\bar{x}_{jt}^{\text{own}} + \bar{x}_{jt}^{\text{other}} + x_{jt}$ does not vary within each year). We therefore use $\bar{x}_{jt}^{\text{own}}$ as excluded instruments to contrast our results with Berry et al. (1995)'s specification. These variables exploit variation in ownership structure across brands and time. Intuititively, this source of variation is more relevant for the price simultaneity problem, than for the endogeneity of market shares. We label those instruments *BLP-IVs*.

In addition, we construct a series of variables that we think are correlated with cost differences across cars. In particular, we exploit variation induced by the differential impact common cost shocks on the relative price of cars with different product attributes. Since most Asian and European cars were produced abroad during this time period, changes in the price oil and in exchange rates lead to changes in the marginal cost of selling cars to the US. Similarly, the relative price of bauxite and iron differentially affect large and small cars.[17] We exploit these sources of variation as follows:

$$w_{jt} = \begin{pmatrix} (\Delta\text{Gas price}_t, \Delta\text{Exchange rate}_t) \times \text{Country of origin}_j \\ \text{Bauxite price/Iron price}_t \times \text{Car size}_{jt} \end{pmatrix}$$

We label those instrumental variables *Cost-IVs*. We use these IVs to estimate the following hedonic price index by OLS:

$$\hat{p}_{jt} = x_{jt}\hat{\pi}_1 + w_{jt}\hat{\pi}_2. \tag{38}$$

This predicted price corresponds to a price index ranking car values based solely on their own observed attributes. We use the results of this first stage regression for two purpose: (1) to recover our estimate of the linear price coefficient β_p, **and** (2) to construct exogenous measures of differentiation.

To construct the *Differentiation IVs*, we extend the approach developed in the Monte-Carlo simulations to account for the multi-dimensional nature of product differentiation. In particular, for each continuous characteristic $y_{jt} \in \left\{\hat{p}_{jt}, \text{HP/WT}_{jt}\right\}$, we construct the histogram of characteristic differences, and the sum of competing product characteristics along six dimensions. More formally, for each nominal difference grid $c^y_{k,t} \in \{c^y_{1,t}, \ldots, c^y_{K,t}\}$:

$$z^y_{jt,k} = \sum_{i \neq j} 1(d^y_{ji,t} < c_{k,t}) \cdot x_{it}$$

where $d^y_{ji,t} = y_{jt} - y_{it}$, and x_{jt} includes continuous and discrete product characteristics.

To account for the change over time in the distribution of characteristics, we define the histogram cutoffs as the within year percentiles of the distribution of $d^x_{ji,t}$, between $1/(K+1), \ldots, K/(K+1)$. In addition, for discrete characteristics such as the A/C and automatic (AT) indicators, we replace the cutoffs with indicator variables equal to one for product sharing the same attribute (e.g., $d^{a/c}_{ij,t} = 0$).

Finally, we construct an approximation of the optimal IV of Chamberlain (1987) similar to the one used in Berry et al. (1999) (see also Reynaert and Verboven (2013)). Recall, that the optimal IV for parameter θ is defined as the conditional expectation of the gradient of the quality assignment function with respect to θ. Berry et al. (1999) approximate this instrument by evaluating the gradients at a preliminary estimate of the parameter value, and replacing each product unobserved product quality by zero. Since this gradient is also function of prices, Berry et al. (1999) replace each product's price by the Bertrand-Nash equilibrium price evaluated at $\xi_{jt} = 0$. Rather than

[17] To account for the fact that car makers are not price takers in the steel or aluminum market, we replace the two input prices with predicted prices using current and lag oil prices.

imposing this type of supply-side assumption, we instead replace prices by the hedonic price index $\hat{p}_{jt}$, as defined in equation 38. Our approximation of the optimal IVs for non-linear parameter θ_k is therefore defines as:

$$z_{jt}^{\theta_k} = \frac{\partial f_{jt}(\xi = 0, p = \hat{p}|\theta = \hat{\theta})}{\partial \theta_k} \tag{39}$$

where $\hat{\theta}$ is a preliminary GMM estimate of the preference parameter vector obtained using either BLP-IVs or Differentiation-IVs.

If IV_{jt} denotes the entire vector instrumental variables, either BLP-IVs, Differentiation-IVs, or Optimal-IVs, the GMM estimator of the non-linear parameter vector is defined as the following nested fixed-point algorithm:

$$\begin{aligned} \min_{\theta \geq 0} \quad & m_n(\theta) A^{-1} m_n(\theta) \\ \text{s.t.} \quad & m_n(\theta) = (\delta(s,\theta) - x\beta(\theta))^T \, \text{IV} \\ & \delta_{jt}(s,\theta) = \sigma_{jt}^{-1}(s_t,\theta) \\ & \beta(\theta) = \left(x' P_w x\right)^{-1} x^T P_w \delta(s,\theta) \end{aligned} \tag{40}$$

where n denotes the number of car $\times$ periods pairs, and P_w is the projection matrix using the Cost-IVs as price instruments.

Notice that our favorite estimator uses only a subset of instrument vector to estimate $\beta(\theta)$, rather than the entire vector IV_{jt}. We find that this approach yields more reasonable estimates of the price elasticity. In particular, using the entire vector differentiation IVs leads to estimates of β_p that are comparable to the OLS estimates, suggesting that some of these variables are not valid excluded variables. As we will see below, this is consistent with our over-identifying tests results. This suggests that the demand system is mis-speciffied. In particular, the model residuals evaluated at the GMM estimate contains measures of product differentiation not captured by our random-coefficient model.

3.2 Estimation results

Table 5 presents the results of a series of specification tests associated with the joint hypothesis that the random coefficients are equal to zero (i.e. IIA tests). Recall, that under the null hypothesis, the market share of each product relative to the outside good, is function only of the produts' own characteristics. Moreover, if the data-generating process does not exhibit this IIA property, the best preditor of quality-assignment function at $\theta = 0$ is a function of the empirical distribution product characteristic differences relative to product each car's characteristics. Since our *Differentiation IVs* are constructed to approximate this empirical distribution, we can test the IIA hypothesis by testing the validity of the exclusion restriction that competitor products' attributes are uncorrelated with the average quality of products, evaluted at $\theta = 0$.

Table 5: Car demand: IIA specification tests results

	(1) Multi.	(2) Price	(3) HP/WT	(4) DFI	(5) FWD	(6) Four cyl.
χ^2	209.9	82.65	73.29	1.737	24.08	10.59
P-Value	1.61e-07	0.00251	0.0176	0.884	0.000210	0.0602
DF	115	50	50	5	5	5

We implement this test by regressing the log share ratios on products' own characteristics, and different combinations of the *Differentiation IVs*, denoted by z_{jt}:

$$\ln s_{jt}/s_{0t} = p_{jt}\beta_p + x_{jt}\beta_x + \mu_j + \tau_t + z_{jt}\gamma + e_{jt}. \tag{41}$$

The null hypothesis that $\gamma = 0$ is tested by estimating β and γ by 2SLS, using our exlcuded cost shifters to instrument for price.

Each column corresponds to a different combination of the Differentiation IVs. Columns (2) through (6) calculates the distribution of competing product attributes separately for each of our five main differentiation variables: hedonic price index, hp/wt, diesel engin (DFI), FWD, and four-cylinders. Column (1) combines all five dimension. To compute the differentiation histograms, we use a 10 uniformly-spaced grid, and interact each indicators with the five characteristics. The three discrete attributes are therefore associated with five instrumental variables (i.e. characteristic of products within the same segment), and the two continuous variables lead to 50 instrumental variables.

The results of the specification tests lead us to easily reject the IIA hypoethesis, especially when we include all five differentiation measures. The tests are less conclusives when looking at each dimension of differentiation indidiually, especially in the case of the diesel and four-cylinder segments. The price index and horse-power measures on the other hand lead to stronger rejections of the IIA hypothesis, at the 1% and 5% significance levels respectively.

Taken together, these results confirm that our Differentiation IVs are jointly able to detect deviations from the IIA property. We interpret this as a evidence that the *Differentiation IVs* are *strong* instruments to identify the random-coefficient parameters. This is because the instruments measure moments of the empirical distribution $F_{jt}(d)$ that are *correlated* with the quality assignment function. The ability of the instruments to explain the quality assignment *away* from the true parameter values, is a necessary condition to identify the model non-linear parameters.

Table 6 present our main set of results. Each entry in Table 6b correspond to a separate single-dimension random-coefficient model. In columns (1) and (2) we compare the parameter estimates obtained with the instrumental variables used in Berry et al. (1995), and with the differentiation IVs defined above. The results are consistent with the Monte-Carlo simulations. The within-firm

Table 6: GMM estimation results for the car application

(a) Single dimension models

	(1) BLP-IV	(2) Diff. IV	(3) BLP (1999)
$\hat{\theta}$			
Price	2.075	1.373	1.254
	(1.36)	(0.25)	(0.3)
HP/WT	2.916	2.611	2.531
	(6.73)	(0.55)	(0.85)
DFI	25.35	2.472	3.2
	(6.83)	(0.55)	(0.23)
FWD	0	1.896	1.823
	–	(0.23)	(0.25)
Four cyl.	0	2.433	2.523
	–	(0.38)	(0.39)
$\hat{\beta}_p$	-3.909	-3.801	-3.796
	(0.75)	(0.77)	(0.65)

(b) Multi dimension models

	Diff. IVs Est.	Diff. IVs S.E.	BLP (1999) Est.	BLP (1999) S.E.
$\hat{\theta}$				
Price	1.11	0.29	1.12	0.31
HP/WT	1.47	0.43	1.53	0.48
DFI	1.21	0.73	1.01	1.57
FWD	1.34	0.18	1.04	0.17
$\hat{\beta}$				
Price	-16.04	0.75	-15.48	0.40
HP/WT	-0.03	0.35	-0.39	0.22
DPM	-0.14	0.20	-0.16	0.19
Car Size	1.38	0.73	1.79	0.50
DFI	-0.36	0.19	0.18	0.14
FWD	0.05	0.08	0.06	0.07

summation of product characteristics lead to very imprecise results, similar to the Market-IVs discussed in the Monte-Carlo simulations. This is not surprising since those variables only vary at the market/firm level, and are uncorrelated with the degree of differentiation of products.

In column (3), we report the results of each model estimated separately using an approximation of the optimal IV of Chamberlain (1987) proposed by Berry et al. (1999) (see also Reynaert and Verboven (2013)). For each specification we use the first-column results as starting values for $\hat{\theta}$.[18] The use of those instruments successfully eliminate the weakness problem found in column (1).

Importantly, parameter estimates and standard-errors are nearly equivalent to the ones found with the Differentiation IVs. This is not surprising, since our main theoretical result suggest that these instruments contain the same information as the optimal IV discussed in Chamberlain (1987). More specifically, we can use Theorem **??** to show that the conditional expectation of the derivative of the quality assignment is an unknown function of the distribution of the characteristic differences:

$$E\left(\frac{\partial \xi_j(\theta)}{\partial \theta_k}\Big| X\right) = E\left(\frac{\partial \xi_j(\theta)}{\partial \theta_k}\Big| F_j(d)\right) = g_k(F_j(d)) \tag{42}$$

Table 6b confirms this results in the multi-dimensional case. In this specification, we estimate a model with four normally distributed random coefficients. The parameters are very precisely estimated for three of the four variables, and are comparable across both sets of of instruments. Notice that the optimal IV specification is estimated using column (1) as starting values, rather than using the market-level IVs results. Therefore, we do not find efficiency gains associated with using the two-stage approach proposed by Berry et al. (1999)

[18]The optimal IV approximation relies on evaluating the derivative of the quality assignment at $\xi_j = 0$. Since this derivative is zero at $\theta = 0$, we use starting values equal to $\min(0.1, \hat{\theta})$ to construct the moment conditions.

A Proof

We make the following regularity assumptions to prove the result.

Assumption 1. *(i) For some $j \in \{1, \ldots, J\}$ we have that $F_b^0 \neq \tilde{F}_b$ implies $\xi_j\left(\cdot; F_b^0\right) \neq \xi_j\left(\cdot; \tilde{F}_b\right)$ (ii) There exists a characteristic x_{j1} that does not admit a random coefficient, i.e., $b_{i1} = \beta_1$*

The condition (i) is a weak regularity conditions that is clearly a minimal necessary conditions for identification. If it fails then the moment restriction (5) would clearly have no identifying power to discriminate between $F_b^0 \neq \tilde{F}_b$. The condition (ii) is a restriction that Berry and Haile (2014) use in a more general preference settings and is sensible for many applications. We adopt it here for convenience, but could potentially relax it (as the next section will make more clear) given that we are working in a more specialized environment.

Part (ii) of Assumption 1 implies that $v_{i1} = 0$. We can partition market structure into $X = \left(X_1, X_{(2)}\right)$ where $X_1 = (x_{j1}, \ldots, x_{J1}) \in \mathbb{R}^J$ are the market structure variables that do not admit random coefficients, and $X_{(2)} = \left(x_{1(2)}, \ldots x_{J(2)}\right) \in \mathbb{R}^{J(K-1)}$ for $x_{j(2)} = (x_{j2}, \ldots, x_{JK})$ are the market structure variables that do potentially admit random coefficients. Then it is clear that for any F_v^0 the quality assignment will be such that $h_j\left(s, X; F_v^0\right) = h_j\left(s, X_{(2)}; F_v^0\right)$. However $g_j(X)$ will nevertheless generally depend upon the entire market structure X (because markets shares s will still depend upon X_1). Thus we have JK potential instruments X for JK potential arguments $\left(s, X_{(2)}\right)$ in h_j. Assuming that the instruments X are *complete* for $\left(s, X_{(2)}\right)$ in the sense of **?**), then the observational equivalence between two models F_v^0 and $\tilde{F}_v$ would imply $g_j(X) =$

$$
\begin{aligned}
E\left[h_j\left(s, X; F_v^0\right) \mid X\right] &= E\left[h_j\left(s, X; \tilde{F}_v\right)\right] \iff \\
E\left[h_j\left(s, X; F_v^0\right) - h_j\left(s, X; \tilde{F}_v\right) \mid X\right] &= 0 \iff \\
h_j\left(s, X; F_v^0\right) &= h_j\left(s, X; \tilde{F}_v\right) \iff \\
F_v^0 &= \tilde{F}_v
\end{aligned}
$$

where the third line follows from the completeness assumption, and the fourth line follows from (i) of Assumption (1).

B Inverse demand representation

Consider the following quasi-linear model:

$$u_{ijt} = \tilde{\delta}_{jt} - \alpha p_{jt} + \sum_k \sigma_k v_{ik} x_{jt,k} + \varepsilon_{ijt} = \delta_{jt} + \sum_k \sigma_k v_{ik} x_{jt,k} + \varepsilon_{ijt} \tag{43}$$

The demand function is given by:

$$\sigma_{jt}(\boldsymbol{\delta}_t, \boldsymbol{X}_t; F_v) = \int \frac{\exp(\delta_{jt} + \sum_k \sigma_k v_{ik} x_{jt,k})}{1 + \sum_{j'=1}^{J_t} \exp(\delta_{j't} + \sum_k \sigma_k v_{ik} x_{j't,k})} dF_v \tag{44}$$

The inverse-demand function is given by:

$$p_{jt} = \boldsymbol{x}_{jt}\boldsymbol{\beta} - \frac{1}{\alpha}\sigma_{jt}^{-1}(\boldsymbol{X}_t, \boldsymbol{s}_t; F_v) = \boldsymbol{x}_{jt}\boldsymbol{\beta} - \frac{1}{\alpha}h_{jt}(\boldsymbol{X}_t, \boldsymbol{s}_t; F_v) + \xi_{jt} \tag{45}$$

Consider the following minimum distance estimator:

1. *Excluded instruments:* $z_{jt} = H(\boldsymbol{X}_t)$

2. *Hedonic regression:*

$$p_{jt} = \boldsymbol{x}_{jt}\hat{\boldsymbol{\gamma}}_{\boldsymbol{x}} + \boldsymbol{z}_{jt}\hat{\boldsymbol{\gamma}}_{\boldsymbol{z}} + \mu_t + \text{Residual}$$

3. *Willingness to pay regression:* For a given guess of the preference parameters $\boldsymbol{\theta}$

$$E\left(\boldsymbol{x}_{jt}\boldsymbol{\beta} - \frac{1}{\alpha}h_{jt}(\boldsymbol{X}_t, \boldsymbol{s}_t; F_v) + \xi_{jt} \middle| \boldsymbol{X}_t\right) = \boldsymbol{x}_{jt}\boldsymbol{\gamma}_{\boldsymbol{x}}(\boldsymbol{\theta}) + \boldsymbol{z}_{jt}\boldsymbol{\gamma}_{\boldsymbol{z}}(\boldsymbol{\theta}) + \mu_t(\boldsymbol{\theta})$$

4. *Norm:*

$$\min_{\boldsymbol{\theta}} ||\hat{\gamma} - \boldsymbol{\gamma}(\boldsymbol{\theta})||$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\boldsymbol{x}}, \boldsymbol{\gamma}_{\boldsymbol{z}})$.

References

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.

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. and A. Pakes (2002). The pure characteristics demand model. manuscript.

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

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

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

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

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

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*.

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

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

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

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–42.

Metaxoglou, K. and C. R. Knittel (2008, October). Estimation of random coefficient demand models: Challenges, difficulties and warnings. working paper, UCSD.

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

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

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

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.

Rosse, J. N. (1970, Mar.). Estimating cost function parameters without using cost data: Illustrated methodology. *Econometrica: Journal of the Econometric Society 38*(2), 256–275.

Staiger, D. and J. H. Stock (1997). Instrumental variable regression with weak instruments. *Econometrica 65*(3), 557–586.

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