Measuring Substitution Patterns in Differentiated-Products Industries*

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

[1]Microsoft & University of Pennsylvania
[2]University of Wisconsin-Madison & NBER

October 8, 2019

Abstract

We study the estimation of substitution patterns within the discrete choice framework developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995). Our objective, is to illustrate the consequences of using weak instruments in this non-linear GMM context, and propose a new class of instruments that can be used to estimate a large family of models with aggregate data. We argue that relevant instruments should reflect the (exogenous) degree of differentiation of each product in a market (*Differentiation IVs*), and provide a series of examples to illustrate the performance of simple instrument functions.

JEL Codes: L13, C35, C36.

*First version: January 18th, 2015. Correspondence: Amit Gandhi, akgandhi@sas.upenn.edu; Jean-François Houde, houde@wisc.edu. This research has benefited from the financial support of the NSF (SES-1530788). We thank the many seminar participants who have provided feedback on this paper. The paper also greatly benefited from discussions with Nikhil Agarwal, Steve Berry, Chris Conlon, Robert Clark, Joachim Freyberger, Paul Greico, Panle Jia Barwick, Robin Lee, Aviv Nevo, Ariel Pakes, Amil Petrin, Katja Seim, and Chris Sullivan.

Introduction

The extent to which competing products are substitutable is central to empirical Industrial Organization (IO) because it informs about the magnitude of market power and consumer welfare in differentiated-product industries. The econometric framework proposed by Berry (1994) and Berry, Levinsohn, and Pakes (1995) is the leading approach for estimating demand in this context, and is increasingly popular as a revealed-preference method to measure quality and value-added in other empirical microeconomics fields.[1] This class of models can approximate very rich substitution patterns by relaxing the *Independence of Irrelevant Alternatives* (IIA) assumption underlying logit/CES type demand structures, while at the same time also accounting for the presence of product-level unobservable attributes (to the econometrician). This flexibility comes at a cost however, since the introduction of non-IIA preferences, either via random-coefficients or a nesting structure, creates a simultaneity problem associated with the joint determination of market shares and unobserved attributes. Importantly, this simultaneity exists wether or not prices (or other characteristics) are also endogenously determined, and has received relatively little attention in the literature.[2]

The endogeneity problem created by market shares is especially challenging for applied work because the parameters that control the substitution patterns enter the demand model in a non-linear fashion. This results in a non-linear GMM estimator, which can be notoriously sensitive to the presence of weak identification; a problem that is difficult to diagnose (e.g. Stock and Wright (2000)). A review of the empirical literature suggests that weak-instruments is potentially a pervasive problem. For instance, there are very few direct applications (known to us) that have found statistically and/or economically significant departures from restrictions.[3] In addition, commonly used moment conditions often lead to numerical optimization problems; another symptom of weak identification in non-linear models (e.g. Metaxoglou and Knittel (2014) and Dube, Fox, and Su (2012)). A related challenge is that non-linear GMM problems do not have a classic "reduced-form" regression (or "first-stage"). This makes it difficult to characterize the fundamental variation in the data that would strongly identify substitution patterns. See Angrist and Pischke (2010) for an articulation of this criticism.

Our goal in this paper is to provide a new class of instruments that are practical to construct and circumvent the weak identification problem in demand models with flexible substitution pat-

[1] Examples include models of residential and school sorting (e.g. Bayer, Ferreira, and McMillan 2007, Nielson 2017), and models of adverse selection in insurance markets (e.g. Starc 2014).

[2] 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. See Berry and Haile (2016) for a review.

[3] For instance, it is common to impose additional "cross-equation" restrictions originating from equilibrium supply assumptions (see Berry et al. 1995, Berry, Levinsohn, and Pakes 1999, Eizenberg 2014), micro moments (see Petrin (2002) and Berry, Levinsohn, and Pakes (2004)), or by using more restrictive models of product differentiation such as the nested-logit or GEV models (e.g. Verboven 1996, Bresnahan, Stern, and Trajtenberg 1997).

terns.[4] To highlight very clearly the source of the problem and our solution, we focus first on the identification of the random-coefficient model with exogenous characteristics. We then illustrate how our approach can be generalized to settings with endogenous prices or characteristics.

The starting point of our analysis is the identification result of Berry and Haile (2014). They showed that the demand function can be non-parametrically identified by imposing a set of conditional moment restrictions relating the mean unobserved quality of each product, and the characteristics of rivals. The "reduced-form" of the model then becomes the conditional expectation of the inverse-demand given the menu of characteristics available. A fundamental challenge when using this result to guide empirical work is that the reduced-form function is a product-specific function of *all* observed product characteristics available in a market, which leads to a curse of dimensionality problem in the number of products. Without further restrictions on the data-generating process or on the structure of the model, it is impossible in general to estimate the reduced-form; that is find instrumental variables that can approximate the conditional-moment restriction arbitrarily well (see Newey 1993, Newey and Powell 2003 and Ai and Chen 2003). Without the reduced form, there is no formal characterization of what constitutes *relevant instruments* for the general random-coefficient model commonly used in the literature. This void is important for the design of a "credible" estimation strategy, since the parameters can be weakly identified by a valid instrument function (even in large sample), despite the fact the model is non-parametrically identified.

Our main theoretical contribution is to provide such a characterization by showing that the curse of dimensionality can be solved using implicit restrictions that the demand structure places on the reduced-form. In particular we show that the reduced-form is a *vector symmetric function* of the distribution observed characteristics *differences* between a given product and the other products available in the same market. This property is rooted in the symmetry of the underlying demand function, which is valid in any linear random-utility model with linear preferences.

This result has important implications: an approximation to the reduced-form can be obtained using basis functions that summarize the distribution of characteristic differences (i.e. exogenous measures of differentiation). Importantly the *number* of basis functions necessary to explain the (unknown) reduced-form is invariant to the number of products in the market. It is this latter result that breaks the curse of dimensionality when constructing relevant instruments.

We show that the instruments should reflect the (exogenous) degree of differentiation of each product in a market. We label this family of instruments *Differentiation IVs*, and provide a series of examples to illustrate how to construct simple basis functions that can be used in various contexts; including settings with correlated random-coefficients, and environments in which the characteristics of consumers differ across markets. Our simulation results show that these instruments, by

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

eliminating the weak IV problem, can improve substantially the precision of the estimates (by a factor of 10 in some cases), and improve the numerical performance and speed of the non-linear optimization algorithms used to estimate the parameters. This good performance is also confirmed by several recent applied papers implementing our instruments (see for instance Miravete, Moral, and Thurk 2018, Coşar, Grieco, and Tintelnot 2018, Singleton 2019, Chaves 2019, Conlon and Gortmaker (2019)).

We then discuss how to adapt our identification strategies to settings in which one or more characteristics is endogenous; exploiting the availability of external instruments (e.g. cost-shifters) or natural experiments. To account for endogenous prices, we combine our main theoretical result with the heuristic approximation suggested by Berry, Levinsohn, and Pakes (1999), and revisited recently by Reynaert and Verboven (2013). We illustrate that both approaches can be quite complementary to one another.

Importantly, this characterization of the reduced-form does not depend on the distribution of the random-coefficient, or on the value of the parameters. Therefore, the same instruments can be used to estimate different models of product differentiation, and to test between alternative specifications. For instance, we use our results to construct a test of the IIA hypothesis, which can be used to measure the strength of the instruments. This is an important advantage of our approach over alternative two-step approximations to the optimal instruments that require obtaining consistent estimates of the parameters, and cannot be used to discriminate between alternative model specifications.

Our paper is not the first to indirectly or directly point out the possibility that aggregate models of product differentiation suffer from a weak identification problem, and propose alternative instrument functions. Recently, Reynaert and Verboven (2013) discussed the loss of efficiency associated with commonly used instruments. The instruments that we propose are also similar to the instruments commonly used to identify nested-logit and spatial differentiation models.[5] We derive an instrument function that can exploit variation in demographic characteristics across markets, similar to the one proposed by Romeo (2010). A key contribution of our paper is to develop a unifying approach to study the estimation of a general family of characteristics models with aggregate data.

The rest of the paper proceeds as follows. In the next section, we formally define the identification and the curse of dimensionality problems in a baseline model with exogenous characteristics. In Section 2 we illustrate and formally diagnose the weak instrument problem. Section 3 include our main sets of results. We first derive our main theoretical result, and illustrate its implication for the construction of *Differentiation IVs*. We then present a series of Monte-Carlo simulation results to analyze the finite-sample properties of the instruments. Finally, in Sections 4 and 5, we extend our model to account for endogenous characteristics, and compare the performance of

[5] See in particular Berry (1994), Bresnahan, Stern, and Trajtenberg (1997), Pinkse, Slade, and Brett (2002), Davis (2006), Thomadsen (2007), and Houde (2012).

our instruments with the optimal IV approximation discussed in Reynaert and Verboven (2013). Appendix A and B include the proof of the main propositions, and computation details related to the Monte-Carlo simulations.

1 Baseline model: Exogenous characteristics

In order to illustrate the instrument choice problem, we consider a special case of the random-utility model considered by Berry et al. (1995), in which product characteristics (including prices) are exogenous.

1.1 Model notation and assumptions

Consider a market t with $J_t + 1$ differentiated-products. Each product j is characterized by a vector of observed (to the econometrician) product characteristics $\boldsymbol{x}_{jt} \in \mathbb{R}^K$ and an unobserved characteristic ξ_{jt}. We will refer to $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{J_t,t})$ as a summary of the observed *market structure* - the entire menu of observed product characteristics available to consumers in market t (i.e. $J_t \times K$ matrix). Similarly, $\boldsymbol{s}_t = \{s_{1t}, \ldots, s_{J_t,t}\}$ is the vector of observed market shares, which is defined such that $1 - \sum_{j=1}^{J_t} s_{jt} = s_{0t}$ is the market share of the "outside" good available to all consumers in market t. We normalize the characteristics of the outside good such that $\boldsymbol{x}_{0t} = 0$ and $\xi_{0t} = 0$.[6]

Following Berry et al. (1995) and Berry and Haile (2014), we assume that the unobserved characteristics of products are independent of the set of characteristics available in each market. Assumption 1 formalizes the main identifying restriction.

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

$$E\left[\xi_{jt} \mid \boldsymbol{x}_t\right] = 0. \tag{1}$$

Our second assumption, refers to the shape of the indirect utility function. We assume that the preference of consumers can be summarized by a linear-in-characteristics random-utility model with a single-index unobserved quality.

Assumption 2. *Each consumer* i *has linear preferences for products* $j = 0, 1, \ldots, J_t$*:*

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

where $\delta_{jt} = \boldsymbol{x}_{jt}'\boldsymbol{\beta} + \xi_{jt}$ *is labelled as the "mean utility" of product* j*,* $\boldsymbol{x}_{jt}^{(2)}$ *is a sub-vector of* $\boldsymbol{x}_{jt}$ *(i.e. non-linear attributes),* $\epsilon_{ijt} \sim T1EV(0,1)$ *is random-utility shock for product* j*, and* $\boldsymbol{\nu}_i = \{\nu_{i1}, \ldots, \nu_{iK_2}\}$ *is the vector of random-coefficients for consumer* i*.*

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

We maintain the mixed-logit parametric functional-form in our simulations below, since it is the workhorse model used in the literature. However, our theoretical results do not depend on this particular distributional assumption, and are relevant for a broader family of characteristic models; including in particular the pure-characteristic and semi-parametric demand models with linear preferences (e.g. Berry and Pakes 2007 and Compiani 2019).

If $F(\boldsymbol{\nu}_i;\boldsymbol{\lambda})$ denotes the joint distribution of the random-coefficient vector, the aggregate demand function for product j can be written as follows:

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

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

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

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

where $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\lambda})$ is the full parameter vector of dimension m. Existence and uniqueness of the inverse demand, $\sigma_j^{-1}(\cdot)$, follows directly from Berry (1994), Berry et al. (1995).[7] When $\nu_{ik}=0$ for all consumers (i.e. multinomial logit), the inverse-demand function is equal to odds-ratio $\ln s_{jt}/s_{0t}$, and the parameters can be estimated by OLS. Otherwise, the market shares of rival products enter the inverse-demand non-linearly, and the least-square estimate of $(\boldsymbol{\lambda},\boldsymbol{\beta})$ will be biased.[8]

Using Theorem 1 from Berry and Haile (2014), it is easy to show that the parameters can instead be identified by imposing the following conditional-moment restrictions (CMR):

$$E\left[\rho_j(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\theta})\middle|\boldsymbol{x}_t\right] = \underbrace{E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\lambda}\right)|\boldsymbol{x}_t\right]}_{\pi_j(\boldsymbol{x}_t;\boldsymbol{\lambda})} - \boldsymbol{x}_t\boldsymbol{\beta} = 0 \quad \text{iff } \boldsymbol{\theta}=\boldsymbol{\theta}^0. \tag{5}$$

We will refer to $\pi_j(\boldsymbol{x}_t;\boldsymbol{\lambda})$ as the *reduced-form* function of the model evaluated at an arbitrary parameter $\boldsymbol{\lambda}$. This function calculates the expectation of product j's inverse-demand, conditional

[7] See also Berry, Gandhi, and Haile (2013) for a general proof that does not rely on the type-1 extreme-value distribution assumption.

[8] To see this, note that the first-order condition of non-linear least-square with respect to $\boldsymbol{\lambda}$ is not satisfied at the true value of the parameters because the market shares enter $\sigma_j^{-1}(\cdot)$:

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

This echoes early discussion on "non-linear IV" methods in Jorgensen and Laffont 1974 and Amemiya 1974.

on the menu of product characteristics available in t. This expectation is taken over the endogenous variables of the model: the vector of market shares.

The standard approach to estimate $\boldsymbol{\theta}$ is to form $L \geq m$ unconditional moment restrictions, consistent with the CMRs defined in equation (5):

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

where $\boldsymbol{z}_{jt} = A_j(\boldsymbol{x}_t)$ denotes an *instrument function* characterizing the menu of characteristics in market t. The instrument vector includes product j's own characteristics, as well as functions of the characteristics of rivals: $A_j(\boldsymbol{x}_t) = \{\boldsymbol{x}_{jt}, A_j^{-x}(\boldsymbol{x}_t)\}$.

How should the instrument function be chosen? Following Newey (1990), Newey and Powell (2003) and Ai and Chen (2003), the role of the instrument function is to approximate the reduced-form function defined by the CMR (see equation 5). In particular, if $A_j(\boldsymbol{x}_t)$ can approximate $\pi_j(\boldsymbol{x}_t, \boldsymbol{\lambda})$ arbitrarily well for all values of $\boldsymbol{\lambda}$, the estimator achieve the efficiency bound defined by Chamberlain (1987)'s optimal instruments. We use this to formally define the relevance of an instrument function.

Definition 1 (Relevance)**.** *An instrument function $A_j(\boldsymbol{x}_t)$ is* **relevant** *if it can approximate the conditional expectation of the inverse-demand function at any $\boldsymbol{\lambda}$ arbitrarily well:*

$$\pi_j(\boldsymbol{x}_t; \boldsymbol{\lambda}) \approx A_j(\boldsymbol{x}_t)\gamma(\boldsymbol{\lambda}) \quad \forall j = 1, \ldots, J_t$$

where $\gamma(\boldsymbol{\lambda})$ is the least-square coefficient vector obtained by regressing the inverse-demand for product j, $\sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda})$, on the instruments $A_j(\boldsymbol{x}_t)$.

If the quality of the approximation is poor, the parameters are weakly identified. Intuitively an instrument function is "strong" if it is a good predictor of the impact of changes to the choice-set of consumers on the inverse-demand of products (i.e. average willingness to pay). This can be thought of as the "the ideal" first-stage of the model.

1.2 IIA Test

The downside of the previous relevance definition is that the reduced-form function depends on an unknown parameter vector $\boldsymbol{\lambda}$, and cannot be empirically assess ex-ante. To get around this problem, we propose a more practical measure of relevance, based on the ability of the instruments to reject the Independence of Irrelevance Alternative (IIA) hypothesis.

With data on individual choices, Hausman and McFadden (1984) propose the following IIA test: estimate the model by including characteristics of rival products in the indirect utility of consumers, and test the exclusion restriction implied by the multinomial logit model. A similar exclusion restriction can be tested using the aggregate inverse-demand function at $\boldsymbol{\lambda} = 0$ (Berry

1994):

$$\begin{aligned}\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\beta}^0,\boldsymbol{\lambda}=0\right) &= \ln s_{jt}/s_{0t} \\ &= \boldsymbol{x}_{jt}\boldsymbol{\beta}^0+\left[\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\beta}^0,\boldsymbol{\lambda}=0\right)-\sigma_j^{-1}\left(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\beta}^0,\boldsymbol{\lambda}^0\right)\right]+\xi_{jt} \\ &= \boldsymbol{x}_{jt}\boldsymbol{\beta}^0+\Delta_j(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\lambda}^0)+\xi_{jt}=\boldsymbol{x}_{jt}\boldsymbol{\beta}+r_{jt}\end{aligned}$$

Importantly, the residual $r_{jt}=\rho_j(\boldsymbol{s}_t,\boldsymbol{x}_t;\boldsymbol{\beta}^0,\boldsymbol{\lambda}=0)$ is correlated with the characteristics of rival products through the $\Delta_j(\cdot)$ function, which measures deviations from IIA in the true model. We construct an "IIA-regression" by taking expectation of shares on both sides, conditional on the menu of product characteristics. This leads to a reduced-form regression relating the log of the odds-ratio to the matrix of product characteristics.

Definition 2 (IIA-test)**.** *If Assumption 1 is valid, the IIA hypothesis can be tested by estimating the following regression:*

$$\begin{aligned}E[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t] &= \boldsymbol{x}_{jt}\boldsymbol{\beta}+E\left[\Delta_j(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\lambda}^0)|\boldsymbol{x}_t\right]+0 \\ &\approx A_j(\boldsymbol{x}_t)\boldsymbol{\gamma}=\boldsymbol{x}_{jt}\boldsymbol{\gamma}_0+A_j^{-x}(\boldsymbol{x}_t)\boldsymbol{\gamma}_1, \qquad (7)\end{aligned}$$

where $A_j^{-x}(\boldsymbol{x}_t)$ is a partition of the full instrument vector including functions summarizing the distribution of products available in market t, excluding product j's own characteristics. The null hypothesis of IIA preferences correspond to: $H_0:\hat{\boldsymbol{\gamma}}_1=0$.

The essence of the test is to quantify the expected deviations from IIA under the true model by estimating a (potentially) mis-specified model.[9] Failure to reject the IIA hypothesis when $\boldsymbol{\lambda}^0\neq 0$ is consistent with weak-identification, since the instrument vector is unable to distinguish between the true model and the "wrong" model (Stock and Wright 2000). To formalize this relationship between weak identification and the IIA regression, in Appendix A.1, we show that the GMM estimator is equivalent to a minimum-distance (MD) estimator that minimizes the distance between the reduced-form IIA regression and the model predicted deviations from IIA. For a given parameter vector $\boldsymbol{\theta}$, the model predicted deviations from IIA are given by the following regression:

$$E\left[\underbrace{\Delta_j(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\lambda})+\boldsymbol{x}_{jt}\boldsymbol{\beta}}_{\equiv y_{jt}(\boldsymbol{\theta})}\Big|\boldsymbol{x}_t\right]\approx A_j(\boldsymbol{x}_t)\boldsymbol{\gamma}(\boldsymbol{\theta})$$

where $\boldsymbol{\gamma}(\boldsymbol{\theta})$ is obtained by projecting $y_{jt}(\boldsymbol{\theta})$ onto the instrument function $A_j(\boldsymbol{x}_t)$. At the true value,

[9] Without controlling for the characteristics of rival products available in market t, the IIA regression suffers from an omitted variable bias. When the instrument is a rich enough control function, in the sense that $E\left[\Delta_j(\boldsymbol{s}_t,\boldsymbol{x}_t^{(2)};\boldsymbol{\lambda}^0)|\boldsymbol{x}_t\right]\approx A_j^{-x}(\boldsymbol{x}_t)\boldsymbol{\gamma}_1$, the omitted variable bias disappears and $\hat{\boldsymbol{\gamma}}_0$ is a consistent estimate of the parameters determining the average willingness to pay of consumers ($\boldsymbol{\beta}^0$).

the difference between $\boldsymbol{\gamma}_n$ and $\boldsymbol{\gamma}(\theta)$ is equal to zero. The distribution of the minimum-distance (and GMM) estimator of θ is the set of parameters that are consistent with the IIA-test statistics estimated from the data. When the instrument function $A_j(\boldsymbol{x}_t)$ fails (or only weakly) rejects the IIA-test, this set can be large, and includes $\boldsymbol{\lambda} = 0$. In other words, the estimated reduced-form is consistent with several models, implying that the minimum-distance moment conditions are weakly satisfied away from the true:

$$E\left(||\boldsymbol{\gamma}_n - \boldsymbol{\gamma}(\boldsymbol{\theta})||\right) : \begin{cases} = 0 & \text{If } \boldsymbol{\theta} = \boldsymbol{\theta}^0, \\ \approx 0 & \text{If } \boldsymbol{\theta} \neq \boldsymbol{\theta}^0. \end{cases}$$

Therefore, an instrument function is "weak" if it cannot detect deviations from IIA using the observed distribution of aggregate market shares and characteristics. Crucially, this definition of relevance can be measured prior to estimating the model.

Of course, the instruments can be "strong" (in the sense of Definition 1) and fail to reject the IIA hypothesis. If that is the case, the researcher should infer that the underlying data-generating process is well approximated by a model with IIA preferences ($\boldsymbol{\lambda}^0 = 0$).

The previous discussion suggests a sequential approach to estimation. First, researchers should evaluate the strength of the proposed instrument function by testing the IIA hypothesis. If the null hypothesis cannot be rejected, the analysis should proceed with the Logit model. Otherwise, the instrument function can be used to estimate a richer model, and test the validity of the over-identifying restrictions. However, as we discuss in Section 3, there exists a (potentially) infinite number of instrument functions to consider, which leads to a curse of dimensionality problem. In Section 3 we use the model to solve this problem, and characterize a class of relevant basis functions that can be used to test the IIA hypothesis and approximate the reduced-form.

2 Illustration of the weak identification problem

The problems of weak identification in linear IV models are well documented (e.g. Stock, Wright, and Yogo (2002)). Similar problems arise in non-linear IV models, but are more difficult to diagnose since the reduced-form of the model depends on the unknown vector of parameters. In this section we illustrate the weak IV problems associated with a commonly used instrument function, and validate the IIA test as a measure of the relevance of the instrument function.

We onsider the following IID random-coefficient model example with **exogenous** characteristics:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + \sum_{k=1}^{K_2} (\beta_{2,k} + \lambda_k \eta_{ik}) \cdot x_{jt,k}^{(2)} + \xi_{jt} + \epsilon_{ijt} \tag{8}$$

where s_{jt} is the observed aggregate market share of product j in market t, and $\eta_{ik} \sim N(0,1)$.

Using the previous notation, $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_{K_2}\}$ denotes the vector of K_2 non-linear parameters. We assume that the number of products is fixed ($J = 15$), and the number of market is equal to $T = 100$.

For this example, we use the sum of rival characteristics as an instrument function:

$$A_j(\boldsymbol{x}_t) \quad = \quad \left\{ \boldsymbol{x}_{jt}, \sum_{j' \neq j}^{J} \boldsymbol{x}_{j',t} \right\} = \left\{ \boldsymbol{x}_{jt}, \text{IV}_{jt}^{sum} \right\}.$$

This is a commonly used instrument in the literature to account for the simultaneous determination of prices (see Berry et al. (1995)). In our context, since characteristics are exogenous, we use this variable as a valid but (potentially) weak moment to identify λ.

Figure 1a illustrates the IIA-test graphically in the single-dimensional model. Each dot represents a product/market combination, and the line corresponds to a linear regression of $\hat{r}_{jt}$ on the instrument.[10] As the figure illustrates, the sum of rival characteristics is uncorrelated with the inverse demand evaluated at $\lambda = 0$, even though the true model exhibits substantial deviations from IIA ($\lambda^0 = 4$). The R^2 and the slope of the regression are both indistinguishable from zero. In other words, the moment conditions are (nearly) satisfied away from the true parameter value ($\lambda^0 = 4$), implying that the model is weakly identified. Importantly, this weak identification problem is not caused by a small sample problem ($N = 1500$). Also the DGP leads to substantial variation in the instrument across markets and products, since we intentionally used a small number of products in our example, $J = 15$. See Armstrong (2016) for discussion of the weak instrument problem for prices when J is large.

The results of $1,000$ Monte-Carlo replications are summarized in Table 1. The first four rows report the average bias and root-mean square error (RMSE) of the estimated log parameters. Note that we estimate the log of λ_k in equation (8), instead of λ_k directly, to account for the strictly positive support of the parameter space. The next four rows report the bias and RMSE of the transformed estimated parameters.

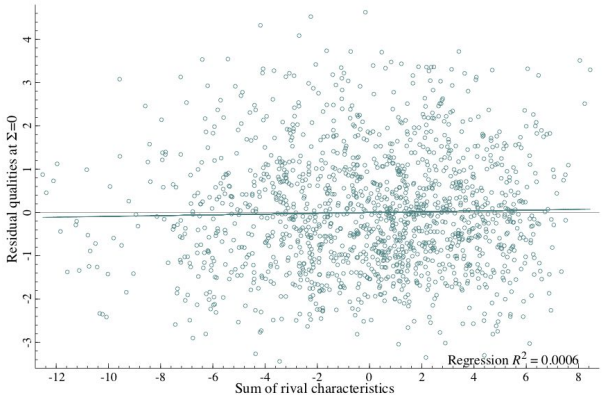
To demonstrate the ability of the IIA-regression to detect the presence of weak instruments, we compare the distribution of the IIA-test with a formal local identification test evaluating the rank of matrix $E\left[\partial \rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right) / \partial \boldsymbol{\theta}^T \cdot \boldsymbol{z}_{jt}\right]$. We use the rank-test proposed by Cragg and Donald (1993) to test the null hypothesis of under-identification under homoskedastic errors. In general, the distribution of the test is sensitive to the weakness of the instruments, since the Jacobian function depends on the *unknown* θ^0. See Wright (2003) for a discussion. This is not an issue in our Monte-Carlo simulations, because we can evaluate the test statistics at the true parameter values, and use techniques common in the linear weak IV literature.[11] The bottom panel of Table

[10] To represent the test graphically we project the instrument onto the product characteristics, and plot the residual on the x-axis.

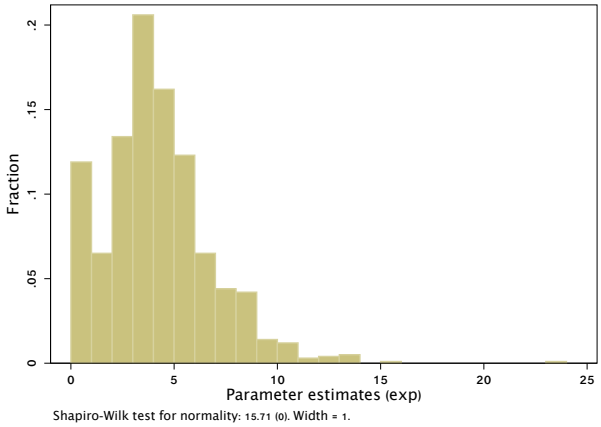
[11] This test can easily be evaluated by testing the rank of $E\left[\partial \rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}^0\right) / \partial \boldsymbol{\theta}^T \cdot \boldsymbol{z}_{jt}\right]$ using standard statistical softwares. For instance, *ranktest* or *ivreg2* in STATA.

Figure 1: IIA test and parameter estimates with weak instruments

(a) IIA Regression



(b) Distribution of $\hat{\lambda}$



1 reports the results of the two tests.

We test the null-hypothesis of IIA preferences by testing the joint null hypothesis that $\hat{\gamma}_1 = 0$. As the figures suggests, we cannot reject the hypothesis of IIA preferences across all four specifications. We reach the same conclusions using the rank-test results. The null hypothesis of under-identification (i.e. rank less than m),, cannot be rejected with probabilities ranging between 60% and 92% on average across the specifications. Note that the p-values with the IIA-test tend to be smaller than the rank-test ones calculated using Stock (2005) critical values. This suggest that the critical values to diagnose weak instruments using the IIA test should be adjusted upward.

Next, we look at the finite-sample performance of the GMM estimator under weak identification. Figure 1b plots the distribution of $\hat{\lambda}_1$ in the $K_2 = 1$ specification. In this specification, 8.4% of $\hat{\lambda}_1$ are estimated to be less than 0.001, which can be interpreted as a corner solution to the GMM optimization problem. This is a robust feature of weak instruments that has been documented by other researchers analyzing the BLP model (e.g. Reynaert and Verboven (2013)).

This "zero problem" is caused by a combination of two factors. First, weak instruments imply that the normal distribution is a poor approximation of the finite-sample distribution of the parameter estimates. This can clearly be seen in Figure 1b; we can easily reject the null hypothesis of normality using Shapiro-Wilk test statistic.[12] In practice, this means that the distribution of the parameter estimates produce frequent outliers. Second, since the parameter space is bounded, outliers on the left-side of the distribution lead to corner solutions, and therefore frequent zeros.

The second panel in Table 1 shows that the dispersion of parameter estimates is still large after transforming the parameter estimates. The RMSEs range from 2.2 to 2.6 across specifications; or more than 50% of the true parameter value (i.e. $\lambda_k = 4$ for all k's). The precision of the estimates

[12]The Shapiro-Wilk test statistic was computed using the empirical distribution of $\ln \hat{\lambda}$.

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

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

Data generating process: $J = 15$ and $T = 100$, $x_{jt}^k \sim N(0,1)$ for $k = 1, \ldots, K$ and $\xi_{jt} \sim N(0,1)$. The parameter values are given by: $\beta_0 = -3, \beta_1 = 1, \beta_2 = 1, \lambda_k = 4$ for all k. Number of simulations: 1,000.

is poor across all four specifications, and remains constant as we increase the complexity of the model.

Another consequence of weak instruments is the presence of numerical optimization problems. To illustrate this point, for each simulated sample, we launched the optimization routine at 10 random starting values (centered around the truth), and use a Nelder-Mead (or Simplex) algorithm to find the local minimum. The indicator variable 1(Local-min) is equal to one if the algorithm converged to more than one solution.

This procedure clearly shows that weak instruments frequently cause multiple local minima. Moreover, the frequency of the problem is increasing with the dimensionality of the parameter space. When $K_2 = 4$, 66% of the samples exhibit multiple minima out of 10 starting values, compared to 19% when $K_2 = 2$. The link between weak instruments and numerical problems is easy to understand. Weak identification implies that the moment conditions are almost satisfied away from the true parameter, which leads to non-convexities in the GMM objective function. This makes it difficult for standard hill-climbing algorithms to find the global minimum when instruments are weak.

The next two rows of Table 1 illustrate the magnitude of the differences between the different

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

3 Instrument relevance and product differentiation

As we discussed in Section 1, an instrument function is relevant if it can approximate the reduced-form of the model arbitrarily well. The choice of instrument function is analogous to choosing a basis-function that can predict the inverse-demand function for all parameter values (Definition 1), as well as deviations from IIA in the (true) inverse-demand function (Definition 2).

In most cases, summarizing the information contained in the conditional moment restrictions to estimate these reduced-form relationships is a daunting task. Recall that the number of exogenous variables $\boldsymbol{x}_t$ is equal to $K \times J_t$, and the number of endogenous variables in the structural equation is equal to J_t. In many applications the number of products is at least as large as the number of markets/periods. This creates a curse of dimensionality problem limiting our ability to approximate the reduced-form; at least without making further restrictions on the shape of the reduced-form function that needs to be approximated.

Formally, a curse of dimensionality exists because the reduced-form of the model is a product-specific function of the entire menu of product characteristics available in the market. As the number of products in each market increases, both the number of arguments and the number of functions to approximate increase.[13] In this context, it is impossible to approximate the reduced-form, because the number of arguments in the approximating functions grows at the same rate as the sample size (a violation of Assumption 3.7 in Ai and Chen (2003)). For the same reason, it is not feasible to approximate Chamberlain (1987)'s optimal instruments using non-parametric regressions, as suggested in Newey (1990).

Therefore, unless the number of products is assumed to be constant and small relative to the number of markets, the number of terms necessary to approximate the function grows exponentially.[14] Intuitively, changes in market structure (from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t'}$) affect each product differentially due to market segmentation, which implies that the expectation of the inverse-demand function

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

[14] Note that this does not affect the identification result in Berry and Haile (2014), since they consider data-generating processes with finitely many products (i.e. $T \to \infty$).

needs to be approximated separately for each product.

This has important implications for empirical work, since there is no formal characterization of what constitutes relevant instruments for the general random-coefficient model commonly used in the literature. As a result the choice of instruments is often poorly justified, which leads to a weak identification problem as we discussed in Section 2.

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

3.1 Demand symmetry and the curse of dimensionality

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

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

$$\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}\right) = f\left(\boldsymbol{\omega}_{jt}; \boldsymbol{\lambda}\right) + C_t(\boldsymbol{\lambda}), \quad j = 1, \ldots, n_t \tag{9}$$

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

The proof can be sketched as follows. We first recognize that the identity of products or the level of product attributes is irrelevant to predict consumers' discrete choice. Therefore, we can abstract from the identity of products by expressing the demand function in terms of characteristics differences relative to product j. Furthermore, rather than normalizing the quality index of the outside good to zero, we rescale the quality index to be between zero and one: $\tau_{jt} = \exp(\delta_{jt}) / \left(1 + \sum_{j't} \exp(\delta_{j't})\right)$ for all $j = 0, \ldots, J_t$. This new normalization has the advantage of treating the outside option symmetrically with respect to the other options, and explains the presence of a market-specific intercept in equation (9).[15] These two normalizations imply that the

[15] The market intercept corresponds to: $C_t(\boldsymbol{\lambda}) = \ln\left(\left(1 - \sum_{j \geq 1} \mathcal{D}^{-1}(\omega_{jt}; \boldsymbol{\lambda})\right)^{-1}\right)$, where $\mathcal{D}(\omega_{jt}; \lambda)$ is the (symmetric) demand function for product j.

demand function for product j is a fully exchangeable function of the structure of the market *relative* to product j: $\boldsymbol{m}_{jt} = \left\{(\boldsymbol{d}^{(2)}_{jt,0}, \tau_{0t}), \ldots, (\boldsymbol{d}^{(2)}_{jt,j-1}, \tau_{j-1,t}), (\boldsymbol{d}^{(2)}_{jt,j+1}, \tau_{j+1,t}), \ldots, (\boldsymbol{d}^{(2)}_{jt,n_t}, \tau_{n_t,t})\right\}$. The inverse mapping associated with this demand representation maintains the same symmetry and anonymity properties.

There are two key implications of Proposition 1. The first is that the inverse-demand function $\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}\right)$ is no longer indexed by product j, once we condition on a vector of state variables $\boldsymbol{\omega}_{jt}$ of the products competing with j in a market.[16] The second implication is that$f(\cdot)$ is a *symmetric* function of the states of the competing products.

To further restrict the shape of the reduced-form function, we impose an additional assumption on the joint distribution of $\{\xi_{1t}, \ldots, \xi_{J_t,t}\}$.

Assumption 3. *The joint distribution of the unobserved quality of products is exchangeable in the identity of products:*

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

for any ordering function $\rho()$.

In economics terms, this assumption implies that the identity of rival products is not important to predict the distribution of unobservable attributes. This does not necessarily rule out the possibility that brands, for instance, are relevant for consumers' decisions. As long as brand or product fixed-effects enter the model linearly (shift the mean attribute), they can be concentrated-out of the residual quality. This assumption is not novel in the literature. It is implicit in much of the prior empirical work, and is discussed explicitly in Berry et al. 1995 (section 5.1).

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

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

$$\pi_j(\boldsymbol{x}_t; \boldsymbol{\lambda}) = g\left(\boldsymbol{d}_{jt}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda})$$

where c_t *is a market specific constant.*

The proof can be sketched as follows. Recall the expectation operator defining the reduced-form function in equation (5) is taken over the market shares vector; which corresponds to the demand functions. Since the demand for each product is symmetric, the density of shares can be re-written as a function of the entire vector of characteristics differences and the joint density of unobservable quality ξ_{jt}. This involves re-ordering the vector of characteristic differences to predict the marginal

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

distribution of each product's market share, and does not require knowing the identity of each individual product (under Assumption 3). This establishes that the expectation of the inverse-demand is a symmetric function of the matrix $\boldsymbol{d}_{jt}$, because the joint distribution of market shares and the integrand itself are symmetric functions of characteristic differences.

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

$$g(\boldsymbol{d}_{jt};\boldsymbol{\lambda}) \approx \sum_{j'\neq j} \gamma_{j'} d_{jt,j'} = \gamma_1 \cdot \left(\sum_{j'\neq j} d_{jt,j'}\right)$$

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

$$g(\boldsymbol{d}_{jt};\boldsymbol{\lambda}) \approx \sum_{j'\neq j}\sum_{k\neq j} \gamma_{j',k} d_{jt,k} d_{jt,j'} = \gamma_1 \cdot \left(\sum_{j'\neq j} d_{jt,j'}\right) + \gamma_2 \cdot \left(\sum_{j'\neq j} \left(d_{jt,j'}\right)^2\right) + \gamma_3 \cdot \left(\sum_{j'\neq j} d_{jt,j'}\right)^2$$

The symmetry property restricts the number of basis-functions to at most three. The first and last terms exhibit very little variation across products. Moreover, conditional on $\boldsymbol{x}_{jt}$, the sum of characteristics differences span the same space as the sum of rival characteristics; which is weakly correlated with the inverse-demand function as we discussed in Section 2. We will therefore focus on the sum of square of characteristic differences to construct our instruments.

In summary, Proposition 1 solves the curse of dimensionality in two ways. First, by expressing the state of the industry in differences (rather than in levels), it is no longer necessary to condition on the identity of products to express the inverse-demand function. This allows us too "pool" observations within and across markets since the same inverse-demand equation is used to explain the data on all products (j,t). Second, under Assumption 3, the expectation of the inverse demand is an exchangeable function of the vector of characteristics difference. This implies that the inverse-demand is function of the magnitude of characteristic differences, not the identity of competing products. As the previous example illustrates, this leads to a substantial reduction in the number of basis functions necessary to approximate the reduced-form.

Proposition 1 extends the partial-exchangeability result obtained in Pakes (1994) to reduce the dimensionality of equilibrium strategies in differentiated product markets (e.g. investment and pricing). In particular, Pakes argues that a firm's demand and profit functions are partially exchangeable in its own and rivals' vector of characteristics: $\sigma_j(x_{jt}, x_{-j,t}) = \sigma_j(x_{jt}, x_{\rho(-j),t})$ for

any ordering $\rho()$. Berry et al. 1995 use this result to construct their instruments for prices: characteristics of own and rival products.

While this result certainly alleviates the curse of dimensionality discussed above, it is silent on how x_{jt} and $x_{-j,t}$ should be interacted when constructing the basis functions. In contrast, by expressing the market structure as a matrix of differences, we obtain a fully exchangeable function. This property is commonly used to alleviate the curse of dimensionality associated with the computation of Markov Perfect Equilibrium (see Doraszelski and Pakes (2007) for a survey of this literature). Farias, Saure, and Weintraub (2012) discusses various moment-based approximation functions that exploit this property. See also Altonji and Matzkin (2005) for a related use of symmetric functions for the estimation of non-separable models.

Finally, the following two corollaries establish a direct connection between product differentiation, optimal instruments, and the IIA test introduced in Section 1. First, a direct implication of Proposition 2 is that the optimal instruments introduced by Amemiya (1977) and Chamberlain (1987) can be written as symmetric functions of the distribution of characteristic differences relative to product j. This is because the symmetry property of the expectation operator used to construct the reduced-form carries over in the construction of the conditional expectation of the Jacobian.

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

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

where $c_{t,k}$ is a market-specific constant.

This implies that it is feasible to find basis-functions that can approximate the optimal instruments, while avoiding the curse of dimensionality problem. The same basis-functions used to approximate the reduced-form can then either be used directly as instruments (as suggested by Berry et al. 1995), or as a two-step procedure to construct the optimal instruments following Newey (1990).

Second, the conditional expectation of the inverse-demand under logit can be written as a symmetric function of the distribution of characteristic differences. This corresponds to a non-parametric IIA regression defined in Section 2. Corollary 2 formalizes this result.

Corollary 2. *If the distribution of $\{\xi_{1t}, \ldots, \xi_{n_t,t}\}$ is exchangeable, the IIA regression can be written as a symmetric function of the matrix of characteristic differences:*

$$E\left[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t\right] = \boldsymbol{x}_{jt}\boldsymbol{\beta} + E\left[\Delta_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^0)|\boldsymbol{x}_t\right] = \boldsymbol{x}_{jt}\gamma_0 + h(\boldsymbol{d}_{jt}) + h_t^0 \tag{10}$$

where h_t^0 is a market-specific intercept.

The key implication of this corollary is that the IIA hypothesis can be tested by regressing the log-share differences on functions summarizing the distribution of characteristics differences. Since the predicted value of this regression measure the conditional expectation of the difference in the inverse-demand function relative to Logit, this test can be used to measure the relevance of a candidate instrument function. This is useful both as a specification test, and as a method to select relevant instrument functions without having to compute the inverse-demand function explicitly.

3.2 Differentiation IVs

Recall that a relevant instrument function is a set of basis-functions that can approximate the reduced-form function. The main implication of Proposition 2 is that a relevant instrument function for $\boldsymbol{\lambda}$ correspond to a finite number of moments characterizing the empirical distribution of characteristic differences relative to product j in market t. Since these functions measure the degree of differentiation, we label them *Differentiation IVs*. To fix ideas, we consider a few examples that we will use in the numerical examples below.

Our first example is based on the idea of constructing an instrument function using the leading terms of a second-order symmetric polynomial basis function (focussing only on the binary interaction terms):

$$A_j(\boldsymbol{x}_t) = \begin{cases} \boldsymbol{x}_{jt} & \text{Own characteristics} \\ \sum_{j' \neq j} \left(d^k_{jt,j'} \right)^2, \quad \forall k & \text{Isolation of product } j \text{ along dimension } k \\ \sum_{j' \neq j} d^k_{jt,j'} \times d^l_{jt,'}, \quad \forall k \neq l & \text{Interaction between dimension } k \text{ and } l \end{cases} \tag{11}$$

where $d^k_{jt,j'} = x_{j't,k} - x_{jt,k}$ measures the difference between product j and j' along dimension k. These functions have an economic interpretation in IO. The sum of square of characteristic differences is a continuous measures of product isolation proportional to the Euclidian distance of product j along each. The interaction terms for which $l \neq k$, capture the covariance between two dimensions of differentiation.

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

This feature suggests the following instrument vector:

$$A_j(\boldsymbol{x}_t) = \begin{cases} \boldsymbol{x}_{jt} & & \text{Own characteristics} \\ \sum_{j' \neq j} 1\left(|d^l_{jt,j'}| < \kappa^k\right), & \forall k & \text{Isolation of product } j \text{ along dimension } k \\ \sum_{j' \neq j} 1\left(|d^l_{jt,j'}| < \kappa^k\right) \times d^l_{jt,'}, & \forall k \neq l & \text{Interaction between dimension } k \text{ and } l \end{cases} \quad (12)$$

where κ^k is a proximity threshold (e.g. standard-deviation of $x_{jt,k}$ across all markets). The second element measures the number of "close-by" rivals along each dimension of differentiation. The interaction of the indicator function with $\boldsymbol{d}_{jt,j'}$ captures the correlation in characteristics between firms that are direct competitors. When characteristics are discrete, the indicator variables can be replaced by $1(d^k_{jt,j'} = 0)$; which can be thought of as a product-segment indicator.

The two formulations of the Differentiation IVs in equations (11) and (12) can include a large number of terms depending on the number of characteristics. In general, it is advisable to select a subset based on the amount of variation across products and/or markets. For instance, it is common for some product characteristics to exhibit very little variation across markets. In Nevo (2001), the non-linear characteristics vary only at the product level (i.e. $\boldsymbol{x}^{(2)}_{jt} = \boldsymbol{x}^{(2)}_{j}$ $j = 1, \ldots, 25$), while prices vary both at the product and the market level. Assuming for simplicity that prices are exogenous (or that the researchers have a valid cost shifter such as the Hausman (1994)'s instruments), one can construct instruments that will be relevant to identify $\boldsymbol{\lambda}$:

$$A_j(\boldsymbol{x}_t) = \left\{ \boldsymbol{x}_{jt}, \sum_{j' \neq j} 1\left(|d^1_{j,j'}| < \kappa^1\right) d^p_{jt,j'}, \ldots, \sum_{j' \neq j} 1\left(|d^K_{j,j'}| < \kappa^K\right) d^p_{jt,j'} \right\}. \quad (13)$$

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

How does this differ from the existing literature? Interestingly, the basis function for the first-order polynomial formulation corresponds to the suggestion in Berry et al. 1995 of using the sum of product characteristics as instruments. The logic of using exogenous measures of differentiation has been used in other settings. However, the relevance of exogenous measures of differentiation is most often justified by their ability to predict prices, rather than to identify the non-linear parameters. There exist two important exceptions: the nested-logit model (e.g. Berry 1994, Bresnahan, Stern, and Trajtenberg (1997)), and models of spatial differentiation (e.g. Pinkse, Slade, and Brett 2002, Davis 2006, Thomadsen 2007, Houde 2012, Singleton 2019). In both literatures, the standard instruments correspond to different versions of the proximity measures described in equation (12). From this perspective, an important contribution of our approach is to formally show that the intuition developed in these prior literatures remains relevant in the more general random-coefficient model.

3.3 Extension: Demographics moments

A restrictive assumption imbedded in the derivation of the random-coefficient model above is that the distribution of consumer preferences is common across markets. When the density function $\phi_t(\cdot|\boldsymbol{\lambda})$ is indexed by t, for instance due to variation in demographic characteristics, the reduced-form function becomes a market-specific function of the distribution of characteristic differences. There are two ways of accounting for this.

First, one could specify separate moment conditions for each market. This approach exploits the large number of products within each market. Two important requirements are that the product characteristics vary significantly across markets and that the number of products per market is large enough.

When this is not feasible, demographic characteristics can be added to the instrument vector in order to pool moments across markets.[17] To see how this source of variation can be combined with the Differentiation IVs introduced above, consider the following single dimension example similar to Nevo (2001) or Petrin (2002):

$$u_{ijt} = \delta_{jt} + b_{it}x_{jt}^{(2)} + \varepsilon_{ijt}, \quad b_{it} = \lambda_y y_{it} + \nu_i,$$
$$\text{Such that, } y_{it} = m_t + \text{sd}_t e_i, \quad e_i \sim \phi_e(\cdot).$$

In this model, the random coefficient b_{it} is composed of a demographic component y_{it} that is distributed according to (known) distribution, and a residual component ν_i that is normally distributed with mean zero and variance λ_ν^2. Importantly, we assume that the distribution of y_{it} can be well approximated by an affine transformation of random variable e_i, which has common density function $\phi_e(\cdot)$ across markets. In other words, differences across markets can be summarized by the location and spread parameters of y_{it}: m_t and sd_t.

Under this new parametrization, we can use directly Proposition 2 to write the reduced-form of the model as a symmetric function:

$$\pi_{jt}(\boldsymbol{x}_t; \boldsymbol{\lambda}) = g\left(\boldsymbol{d}_{jt}, \text{sd}_t \cdot \boldsymbol{d}_{jt}^{(2)}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda}) \tag{14}$$

See Appendix A.4 for a detailed derivation.

When demographic characteristics are incorporated in the model in this fashion, the reduced-form is a symmetric function of the distribution of characteristics difference, and moments of the distribution of demographic characteristics interacted with differentiation. Moreover, the linear index includes an additional characteristic: $m_t \cdot x_{jt}^{(2)}$. This interaction between "own" characteristic and the market-average of y_{it} can be used as excluded instruments. Therefore, in this example, the instrument vector should include the full vector of products' own characteristics (including

[17] While previous papers have used this type of instruments, they are typically motivated as predictors of markups or choice-sets (see for instance Waldfogel (2003), Gentzkow and Shapiro (2010) and Fan (2013)).

$m_t \cdot x_{jt}^{(2)}$), as well as moments of the distribution of characteristics differences interacted with the standard-deviation of y_{it} in market t.

The argument can easily be extended to multiple dimensions of heterogeneity, as long as the distribution of demographic characteristics can be standardized across markets. For more general families of distributions, we rely on the following heuristic approximation of the reduced-form:

$$\pi_{jt}(\boldsymbol{x}_t; \boldsymbol{\lambda}) \approx g\left(\boldsymbol{d}_{jt}, \mathbf{M}_t \otimes \boldsymbol{d}_{jt}^{(2)}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda}) \tag{15}$$

where $\mathbf{M}_t$ is now a vector of moments characterizing the joint distribution of demographic characteristics in market t.

The key insight of this transformation is that demographic characteristics should enter the instrument vector as interaction terms with other measures differentiation, rather than as stand-alone variables.[18] Heuristically, as long the distribution of demographics can be well approximated by an affine transformation of a vector of moments $\mathbf{M}_t$, including those additional state variables will provide a good approximation to the reduced-form.

We can construct instrument variables that identify separately the two sources of heterogeneity using basis functions similar to the ones discussed above. For instance, the quadratic basis function becomes:

$$A_j(\boldsymbol{x}_t) = \left\{ \boldsymbol{x}_{jt}, m_t \cdot \boldsymbol{x}_{jt}^{(2)}, \sum_{j' \neq j} d_{jt,j'}^1 \times d_{jt,j'}^l, \ldots, \sum_{j' \neq j} d_{jt,j'}^K \times d_{jt,j'}^l, \sum_{j' \neq j} \mathbf{M}_t \otimes \left(d_{jt,j'}^l \times d_{jt,j'}^{(2)} \right) \right\}_{l=1,\ldots,K}.$$

Focussing on the quadratic term (i.e. $d_{jt,j'}^l = d_{jt,j'}^{(2)}$), the added instruments capture how product differentiation asymmetrically impacts the inverse-demand of product j depending on the distribution of demographic attributes of consumers. See Miravete, Seim, and Thurk (2018) for an example of this type of instrument function.

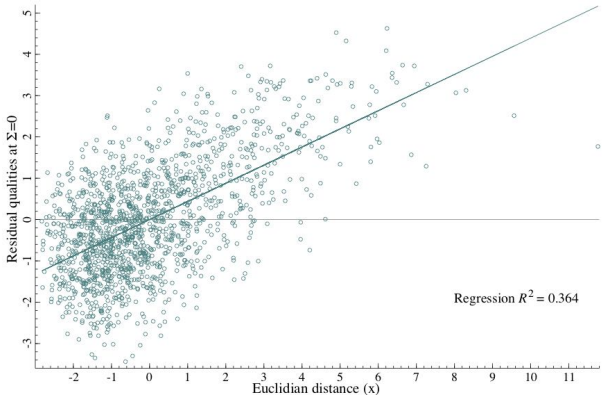
3.4 Monte-Carlo simulations

In this section, we analyze the finite sample properties of the Differentiation IVs described in the previous section. We consider two random-coefficients models with exogenous characteristics: (i) independent random-coefficients: and (ii) correlated random-coefficients. Appendix B provides more details on the data-generating process and the numerical algorithm used for estimation. We use an iterative nested-fixed-point Gauss-Newton Regression (GNR) algorithm, combined with a Newton-Raphson non-linear equation solver, to solve the non-linear GMM problem. This procedure is very robust in settings with strong instruments, and is substantially faster than alternative

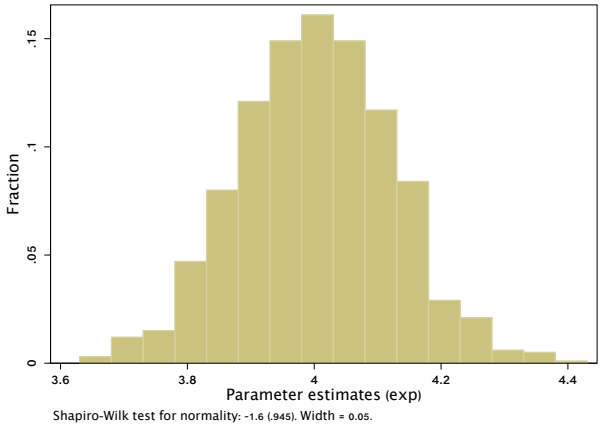
[18] See Romeo (2010) for a similar argument and simulation results showing the importance of accounting for interactions between product characteristics and the mean of demographic attributes in the instrument vector.

Figure 2: IIA test and parameter estimates with strong instruments

(a) IIA regression



(b) Distribution of $\hat{\lambda}$



optimization algorithms. We provide a pseudo-code description of our approach in Appendix B.2. We also provide sample Phython and Ox codes on our website.[19]

Independent random-coefficients

We start by revisiting the numerical example discussed in Section 2:

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

where $\eta_{ik} \sim N(0,1)$ and $\epsilon_{ijt} \sim \text{T1EV}(0,1)$.

Before discussing the results of the Monte-Carlo simulations, we illustrate graphically the relationship between differentiation and the IIA regression introduced in Section 2. Figure 2a illustrates the correlation between the residual function at $\lambda = 0$ (Logit) and the *Euclidian* distance of product j, for the single-dimension heterogeneity model.[20] Unlike the sum of rival characteristics, the Euclidian distance is strongly correlated with the model residual evaluated at $\lambda = 0$; the R^2 of the regression removing the effect of $\boldsymbol{x}_{jt}$ is over 0.35 (compared to 0.0006 in Figure 1a). The Euclidian distance is therefore a good predictor of the inverse-demand function away from the true parameter.

The sign of the correlation between the differentiation and $\rho_j(\boldsymbol{s}_t, \boldsymbol{x}; \lambda = 0)$ reflects the elasticity of substitution at the true parameter λ^0. Since the data is generated by a model with non-IIA preferences, products located in denser areas of the product space have relatively small market shares. The inverse demand evaluated at $\lambda = 0$ rationalizes this feature by assigning high quality

[19] The codes are available here: https://jfhoude.wiscweb.wisc.edu/research-in-progress/

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

Table 2: Simulation results for the exogenous characteristic model with Differentiation IVs

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

Data generating process: $J = 15$ and $T = 100$, $x_{jt}^k \sim N(0,1)$ for $k = 1, \ldots, K$ and $\xi_{jt} \sim N(0,1)$. The parameter values are given by: $\beta_0 = -3, \beta_1 = 1, \beta_2 = 1, \lambda_k = 4$ for all k. Number of simulations: 1,000.

to products that are relatively isolated, and low quality to products with many substitutes. A clear violation of the moment conditions. This positive relationship between differentiation (or distance) and the inverse demand at $\lambda = 0$ is captured by the differentiation instrument used in Figure 2a. Figure 2b confirms that this leads to precise and unbiased parameter estimates. The finite sample distribution of $\hat{\lambda}$ is well approximated by a normal distribution (Shapiro-Wilk p-value = 0.945).

Next we simulate richer models with dimensions of heterogeneity ranging from $K_2 = 2$ to $K_2 = 7$. For each specification, we compare the performance two differentiation IVs:

$$\text{Quadratic Diff IV:} \qquad A_j(\boldsymbol{x}_t) = \left\{ \boldsymbol{x}_{jt}, \sum_{j'} \left(d^1_{jt,j'}\right)^2, \ldots, \sum_{j'} \left(d^K_{jt,j'}\right)^2 \right\}$$

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

where K is the number of characteristics (excluding the intercept), and sd_k is the standard-deviation of $x_{jt,k}$. Table 2 summarizes the small-sample performance for the two IVs across all specifications, and calculates the average asymptotic standard-errors. Tables A1a and A1b in the Appendix summarize the full set of simulation results, including the weak identification and IIA tests, and the local minimum statistics.

Both specifications allow us to reject the null hypothesis of under-identification (rank-test), as well as the IIA hypothesis. In addition, the frequency of local optima is equal to zero across all specifications; meaning that the Newton optimization algorithm always converges to global minimum irrespectively of the starting values. The precision and bias of the parameter estimates are also small across all specifications. The average RMSEs of $\hat{\lambda}_k$ are roughly 17 times smaller with the two instruments defined above, compared with the sum of rival characteristics used in Table 1.

We also find that the loss in precision from adding random-coefficients is very minor. The

average RMSEs increase from 0.03 to 0.05 when we vary the number of random-coefficients from one to seven. This is encouraging since the sample size is fairly small: 15 products $\times$ 100 markets. In addition, the asymptotic standard-errors are nearly identical to the RMSE across all specifications, suggesting that the asymptotic distribution is well approximated by a normal. The proposed instrument functions therefore solve the weak identification problem.

Correlated random-coefficients

Next, we consider a model with correlated random-coefficients:

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

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

To generate the data, we set the diagonal element of $\boldsymbol{\Sigma}$ equal to 4; the same value used in the previous simulations. The covariance terms are chosen such that there is an equal number of positive and negative parameters, equal to either -0.5 or 0.5. See Table 6 in the Appendix.

Note that we estimate Choleski decomposition of $\boldsymbol{\Sigma} = \boldsymbol{C}'\boldsymbol{C}$, rather than $\boldsymbol{\Sigma}$ directly. This allows us to write indirect utility of consumers as a linear function of parameters and K_2 standard-normal random-variables: $\boldsymbol{\nu}_i = \boldsymbol{C}'\boldsymbol{\eta}_i$ where $\boldsymbol{\eta}_i \sim \mathcal{N}(0, I)$. To ensure that $\boldsymbol{\Sigma}$ is positive semidefinite, we constraint the diagonal elements of $\boldsymbol{C}$ to be positive by estimating the log of $C_{k,k}$. Let $\boldsymbol{\lambda}$ denotes the lower-diagonal elements of this transformed matrix.

To construct our instrument function, we use the second-order polynomial form of the Differentiation IVs with additional interaction terms between each characteristics pairs:[21]

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

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

The simulation results are summarized in Table 3. The top panel reports the average estimated parameters (transformed) of the variance-covariance matrix, the middle panel reports the RMSE associated with each parameter, and the bottom panel reports the averages of the IIA-test and

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

Table 3: Simulation results for the correlated random-coefficient model with Differentiation IVs

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

Data generating process: $J = 50$ and $T = 100$, $x_{jt}^k \sim N(0,1)$ for $k = 1, \ldots, K$ and $\xi_{jt} \sim N(0,1)$. The parameter values are given by: $\beta_0 = -3, \beta_1 = 1, \beta_k = 1$ for all k. Table 6 in the Appendix presents the variance-covariance matrix of ν_{ik}. Number of simulations: 1,000.

the Cragg-Donald rank test statistics. Both tests confirm that the instruments are strong, and that the IIA hypothesis is easily rejected. The average bias and RMSE are also small, despite the richness of the model. The differentiation IVs are able to accurately identify both the magnitude and correlation in taste heterogeneity across consumers.

It is worth noting that this specification is substantially richer than any random-coefficient model that has previously been studied with aggregate data by researchers, both in empirical applications and Monte-Carlo simulations. Although we obtain these results in a "controlled" environment, this result confirms the ideas in Berry et al. 1995 and Berry and Haile (2014) that it is feasible to estimate very flexible substitution patterns using aggregate data on market shares and product characteristics.

4 Extension: Endogenous attributes

4.1 Reduced-form and instruments

Incorporating endogenous characteristics, such as prices or advertising, does not fundamentally change the identification problem, but adds an additional simultaneity problem: in equilibrium these characteristics are correlated with the unobserved quality of products (Berry, Levinsohn, and Pakes 1995).

To see how this changes the reduced-form function, consider the following inverse-demand with

endogenous prices:

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

Element k of the state vector $\boldsymbol{\omega}_{jt}$ now includes: $\{s_{kt}, \boldsymbol{d}_{jt,k}^{(2)}, \boldsymbol{d}_{jt,k}^{p}\}$, where $d_{jt,k}^{p}$ is the price differences between product j and k. This inverse demand is generated from a model in which consumers have heterogenous price coefficients (as in Bresnahan (1987) for instance).

As before, $f(\cdot)$ is a symmetric function of the industry state vector $\boldsymbol{\omega}_{jt}$. Although the conditional expectation of this function is also symmetric, the conditional mean restriction used equation (5) to identify the model is no longer satisfied at $\boldsymbol{\theta}^0$:

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

Two broad sources of variation have been proposed in the literature to construct valid price instruments: (i) markup-shifters (e.g. Berry et al. 1995), and (ii) cost-shifters (e.g. Nevo (2001)). Let $\boldsymbol{w}_{jt}$ denotes a vector of relevant and valid price instruments such that:

$$E[\xi_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t] = 0. \tag{16}$$

The challenge in using this restriction to construct instruments, is that the reduced-form of the model cannot be written as a symmetric function of the distribution of characteristics differences $\{\boldsymbol{d}_{jt}, \boldsymbol{d}_{jt}^{w}\}$. To see this, recall that the symmetry of the reduced-form arises from the symmetry of the demand function itself. With endogenous prices, the conditional expectation of the inverse demand is taken with respect to the joint distribution of $(\boldsymbol{s}_t, \boldsymbol{p}_t)$ given $(\boldsymbol{x}_t, \boldsymbol{w}_t)$, which is determined endogenously by the conduct of the industry. Except in special cases such as perfect-competition or single-product Bertrand-Nash, this distribution is not a symmetric function of characteristic differences. This is because the identity/ownership of products plays an important role in determining the distribution of markups.

It is therefore impossible to obtain an exact characterization of the reduced-form without knowing the exact supply relation governing prices. Importantly, this does not mean that it is infeasible to construct valid/relevant instruments. It simply means that we cannot solve the curse of dimensionality problem without relying on a heuristic approximation.

To get around this problem, we rely on the heuristic first proposed by Berry, Levinsohn, and Pakes (1999), and recently reexamined by Reynaert and Verboven (2013). The argument proceeds in two steps.

First, let $\hat{p}_{jt} \approx E(p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t)$ denotes an estimate of the reduced-form pricing equation constructed from observed characteristics. This *exogenous* price measure can be constructed using regressions exploiting random variation from cost and/or ownership shocks (as in Reynaert and

Verboven (2013)), or by solving an equilibrium pricing game after setting $\xi_{jt} = 0$ (as in Berry, Levinsohn, and Pakes (1999)). The choice of the approach is application/data specific. For instance, when using a regression approach, it may be advisable to use flexible functional forms or non-parametric regression techniques to improve the quality of the fit. Since $\hat{p}_{jt}$ is constructed from $(\boldsymbol{x}_{jt}, \boldsymbol{w}_{jt})$, the following conditional moment restriction is satisfied:

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

Second, following Berry, Levinsohn, and Pakes (1999), we use the following heuristic approximation of the reduced-form:

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

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

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

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

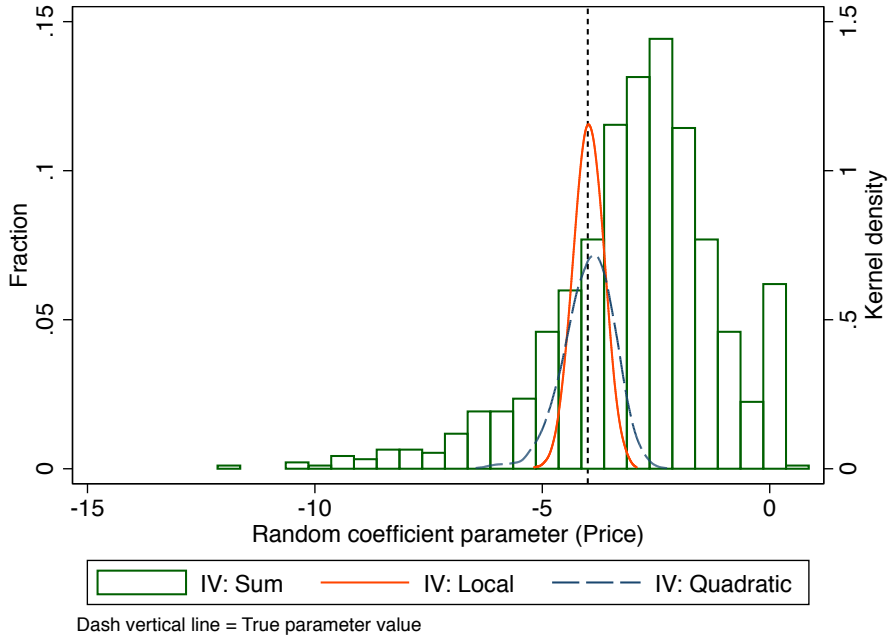
where the vector of characteristic differences $\boldsymbol{d}_{jt,j'}$ is expanded to include (exogenous) price differences $d_{jt,j'}^{\hat{p}}$. Note that $\hat{p}_{jt}$ is included in $A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)$ to instrument for p_{jt}. Similar instruments can be constructed using the sum of square of characteristics differences or the Euclidian distance.

We can use the same heuristic and the above instrument function to test the IIA hypothesis with endogenous prices:

$$\begin{aligned} E[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t] &= \boldsymbol{x}_{jt}\boldsymbol{\beta} + \alpha E[p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t] + E\left[\Delta_j(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^0)|\boldsymbol{x}_t, \boldsymbol{w}_t\right] + 0 \\ &\approx \boldsymbol{x}_{jt}\boldsymbol{\gamma}_0 + \alpha\hat{p}_{jt} + A_j^{-x}(\boldsymbol{x}_t)\boldsymbol{\gamma}_1, \end{aligned}$$

The IIA-test corresponds to the null hypothesis: $H_0 : \hat{\boldsymbol{\gamma}}_1 = 0$. With the "local differentiation" instruments, $A_j^{-x}(\boldsymbol{x}_t)$ includes the number of products with "similar" attributes (including $\hat{p}$), as well as possibly additional interaction terms.

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



4.2 Monte-Carlo simulations

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

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + (\beta_p + \lambda_p \nu_i) \cdot p_{jt} + \xi_{jt} + \epsilon_{jt}, \quad j, = 1, \ldots, 15 \text{ and } t = 1, \ldots, 100. \tag{17}$$

where $\ln \nu_i \sim \mathcal{N}(0, 1)$.[22]

To generate a second simultaneity problem, we generate prices using a Bertrand-Nash pricing game with single-product competitors. Prices are determined by the following vector of first-order conditions:

$$p_{jt}^* = c_{jt} - \sigma_j(\boldsymbol{\delta}_t, \boldsymbol{p}_t^*; \lambda_p) \left[\frac{\partial \sigma_j(\boldsymbol{\delta}_t, \boldsymbol{p}_t^*; \lambda_p)}{\partial p_{jt}^*} \right]^{-1} \tag{18}$$

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

The marginal-cost function is assumed to be constant, and the cost-shock ω_{jt} is observed by the econometrician. We use this variable below to construct a price instrument. The data is generated

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

Table 4: Monte-Carlo simulation results for endogenous price specification

(a) Distribution of parameter estimates

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

(b) Weak identification tests

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

Data generating process: $J = 15$ and $T = 100$, $\beta_0 = 50$, $\beta_x = 2$, $\beta_p = -0.2$ and $\lambda_p = -4$. Number of simulations: 1,000.

by finding a solution to equation (18) for 1000×100 independent markets.[23] This leads to 1,000 simulated panels of market shares and characteristics.

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

$$\hat{p}_{jt} = \hat{\pi}_0 + \hat{\pi}_1 x_{jt}^{(1)} + \hat{\pi}_2 \omega_{jt}. \tag{19}$$

We use a linear functional form for expositional purposes. In applications, richer functional forms might lead to further efficiency gains by improving the fit of the reduced-form regression (see

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

Reynaert and Verboven (2013) for a discussion).

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

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

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

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

The simulation results are reproduced in Figure 4 and Table 4. In addition to the two sets of instruments defined above, we also report the results using the "sum of rival characteristics" in order to illustrate effect weak instruments. The bottom panel reports the results of the weak identification and IIA tests.

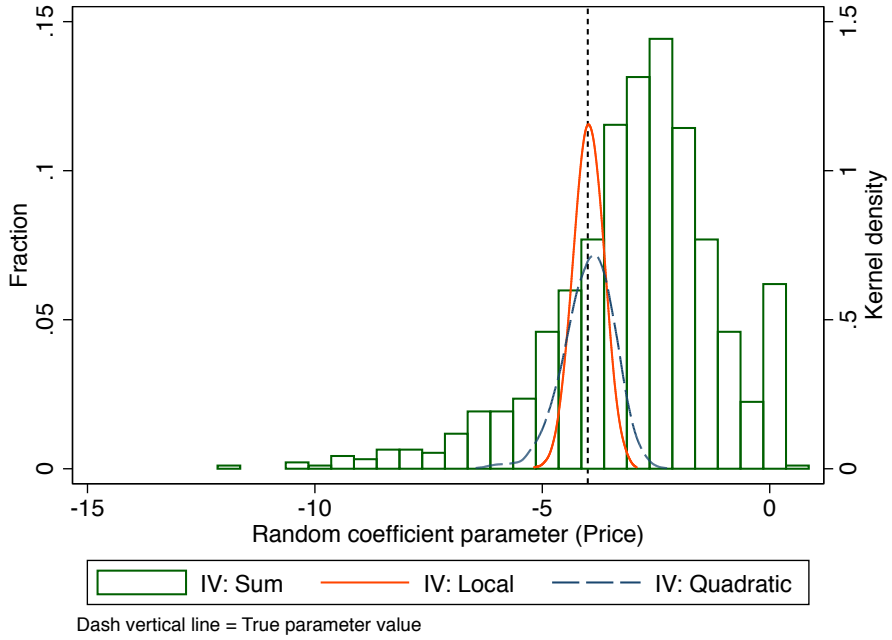
Table 4b confirms the weak identification results obtained in the models with exogenous characteristics. The average test statistics associated the IIA hypothesis and Cragg-Donald rank-test are below standard critical values when using the sum of characteristics as instruments (column 3). In contrast, the sum of square of characteristic difference (column 2) and the number of "local" rival products (column 1) both eliminate the weak identification problem. In this example, the "Local Differentiation IV" tends to perform better than the "Quadratic Differentiation IV". The two measures of weakness, the IIA-test and the Cragg-Donald statistic, are on average roughly 2.5 times larger in column (1) than in column (2). In both specifications, we can reject the null of IIA preferences and under-identification.

To illustrate the dual role of the instruments in this context, Table 4b also report the results of two first-stage F tests: (i) one that simply regresses price and the Jacobian on the exogenous variables, and (ii) one that first "projects-out" the exogenous variation induced by the other endogenous variable before computing the first-stage F test. The second test was proposed by Angrist and Pischke (2009) and Sanderson and Windmeijer (2016) to adjust the standard first-stage tests for cases with multiple endogenous variables.[24]

In our example, the standard F-tests conducted using the sum of rival characteristics incorrectly suggest that weak instruments is not a concern (i.e. 138.94 and 27.85). This is because one of the instrument is very strong (i.e. cost shifter ω_{jt}). However, once we account for the fact that we

[24] We use weak-identification tests designed to test the relevance in linear IV models, by evaluating the Jacobian of the residual function at the true value of the parameters. In practice we obtain similar results when constructing the tests at the GMM estimates instead, but there are no critical values available in the literature for this test. Deriving the limiting distribution of these statistics under weak-identification is beyond the scope of this paper.

Figure 4: Distribution of estimated price random-coefficient parameter for alternative differentiation instruments



have more than one parameters to identify with a single strong instrument, the conditional first stage F-tests are in line with the results of the Cragg-Donald and the IIA tests; both F-tests are significantly below the Stock-Yogo critical values on average.

Table 4a summarizes the distribution of the estimated parameters across the three IV specifications. Looking first at specification (3), we see again that using weak instruments lead to substantial loss in precision and large biases. The RMSE for λ_p is equal 2.10, and the average bias is significantly above zero (1.03). This upward bias is partially offset by a "downward" bias in β_p (i.e. -0.67), but the net effect is positive: weak instruments in this example biases the slope of the demand towards zero.

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

Figure 4 illustrates this point graphically by plotting the distribution of $\hat{\sigma}_p$ for the three specifications. As with the exogenous characteristics, weak instruments lead to a non-Gaussian distribution of the parameters, characterized by large outliers and a mass around zero. The two other distributions are symmetric and bell-shape, centered around the true parameter, and do not exhibit outliers. The comparison between the spread of the distributions with "Quadratic" and "Local"

Differentiation IVs also illustrates the gain in precision from using a stronger set of moments.

4.3 Natural Experiments

An often expressed criticism of the main identifying assumption in Berry et al. 1995, is that firms endogenously choose product characteristics (observed *and* unobserved). This violates Assumption 1 either because of the endogenous selection of products, and/or because of a contemporaneous correlation between ξ_{jt} and the attributes of own and rival products.[25] This invalidates the use of the entire distribution of characteristic differences to identify substitution patterns.

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

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

where $j = 1, \ldots, 15$ indexes products, $m = 1, \ldots, 100$ indexes markets, and $t = 0$ or 1 indexes the pre/post natural experiment periods. In this example, the non-linear characteristic of products, x_{jmt}, measures their location in the product space, and the random-coefficient, ν_i, measures the "ideal" address of consumers. We assume that both variables are uniformly distributed between 0 and 10. The goal is to estimate the travel cost of consumers: λ.

We consider a natural experiment associated with the entry of a new product in each market at location $x^* = 5$ in the post-period (i.e. $t = 1$). Within each market, distance to x^* measures the strength of the "treatment". The characteristics of incumbent products are constant across periods (i.e. $x_{jmt} = x_{jm}$).

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

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

where $\text{ED}_{jm} = \sqrt{\sum_{j'}(x_{jm} - x_{j'm})^2}$ is the Euclidian distance of incumbent product j. The parameter a creates a standard simultaneity problem: products facing close substitutes have higher unobserved quality. Since characteristics are constant across the two periods, this correlation can be absorbed by conditioning on product/market fixed-effects. Assumption 4 formalizes this quasi-experimental design assumption.

[25] See Ciliberto, Murry, and Tamer (2016) for a recent examination of this problem.

Assumption 4. *The change in the unobserved quality of products is mean zero conditional on the observed menu of characteristics and product/market fixed-effects μ_{mt}:*

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

where $\Delta\xi_{jm} = \xi_{jm1} - \xi_{jm0}$ and $\xi_{jm1} = \mu_{jm} + \Delta\xi_{jm}$

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

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

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

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

where n is the number of unique market/product observations. Using this specification, the structural parameters of the model are identified solely from the quasi-experimental variation. In particular, the reduced-form is approximated by a difference-in-difference regression, in which the "control" group is defined as the set of products located relatively far from the exogenous new entrant.

Figure 5 illustrates the ability of this identification strategy to eliminate the simultaneity bias associated with the endogenous location of products. The dash curves correspond to the Kernel density of the parameters estimated using the "difference-in-difference" moment conditions (5a), or the full "Differentiation IVs" moments (5b).[26]

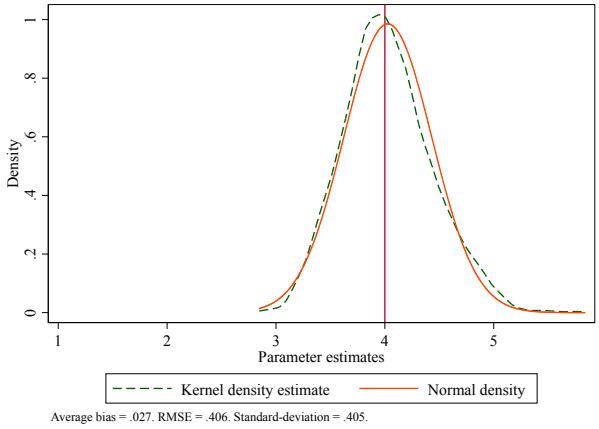
The data generating process is designed so that the correlation between ξ_{jmt} and the Euclidian distance between rival products is $a = -0.25$. As Figure 5b illustrates, this leads to an attenuation bias in the estimate of the travel cost parameter obtained using standard instruments ($\hat{\lambda} \approx 1.89$, compared to $\lambda^0 = 4$). Since products located in "denser" regions of the product space have higher quality, the GMM specification that exploits variation in the distance to all products wrongly infer that consumers have a small disutility from distance. Figure 5b illustrates that the difference-in-difference moment conditions eliminate this bias. The distribution is centered around $\lambda^0 = 4$, and the average bias is less than 1% of the parameter value.

Comparing the two distributions, it is important to note that by exploiting solely the variation created by the entry of a new product, the difference-in-difference GMM estimator is less precise,

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

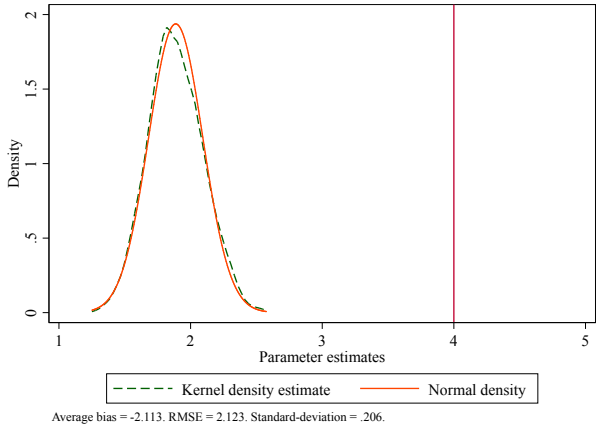
Figure 5: Monte-Carlo simulated distribution of the travel cost parameter estimates with endogenous product locations

(a) Difference-in-Difference Moments



Average bias = .027. RMSE = .406. Standard-deviation = .405.

(b) Full Differentiation Moments



Average bias = -2.113. RMSE = 2.123. Standard-deviation = .206.

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

and the distribution of $\hat{\lambda}$ is less well approximated by the normal density than the specification that uses the larger set of instruments. In Figure 5b the p-value associated with Shapiro-Wilk normal test is 11%, compared to less than 1% in Figure 5a. This suggests that the asymptotic approximation used to conduct inference on λ is less likely to be valid when the model is estimated solely using quasi-experimental variation; therefore requiring larger sample sizes or inference methods that are robust to weak identifications. Alternatively, additional equilibrium restrictions can be used to solve the simultaneity problem (as in Ciliberto, Murry, and Tamer (2016)).

5 Comparison with other approaches

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

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

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

This is very intuitive: Because the asymptotic distribution of $(\boldsymbol{\lambda}, \boldsymbol{\beta})$ is derived from a first-order approximation of the residual function, the most efficient instruments correspond to the best-predictor of the slopes of that function with respect to each of the parameters.[27]

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

$$E\left[\frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big| \boldsymbol{x}_t\right] \approx \frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\xi_{jt}=0, \forall j,t} = \tilde{A}_j(\boldsymbol{x}_{jt}|\boldsymbol{\theta}). \tag{20}$$

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

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

$$E\left[\frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \Big| \boldsymbol{x}_t, \boldsymbol{w}_t\right] \approx \frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}\Big|_{p_{jt}=\hat{p}_{jt}, \xi_{jt}=0, \forall j,t} = \tilde{A}_j(\boldsymbol{x}_{jt}|\boldsymbol{\theta}), \tag{21}$$

where $\hat{p}_{jt} \approx E(p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t)$ is a "reduced-form" model for prices independent of ξ_{jt}.

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

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

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

[27] See Newey (1993) for an illuminating discussion.

Table 5: Optimal IV approximation with alternative initial parameter values

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

Data generating process: $J = 15$ and $T = 100$, $x_{jt}^k \sim N(0,1)$ for $k = 1, \ldots, K$ and $\xi_{jt} \sim N(0,1)$. The parameter values are given by: $\beta_0 = -3, \beta_1 = 1, \beta_k = 1, \lambda = 2$ for all k. Number of simulations: 1,000.

this Hotelling model is a special case of the linear-in-characteristics random-coefficient model.

Table 5 summarizes the results of 1,000 Monte-Carlo replication simulations. The first five rows correspond to different values of the initial parameter used to evaluate the Jacobian. In both specifications, the true value of parameter is $\lambda^0 = 2$. The numbers in bold correspond to GMM results obtained by setting the first-stage parameter equal to the true parameter value . The rest of the rows correspond to different levels of inconsistencies. For the "Normal RC" mode, we consider a grid between 0.5 and 3. For the Hotelling model, we consider grid between -2 (wrong sign) and 4.

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

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

The last row of Table 5 reports the results obtained with the Differentiation IVs. To obtain these results we combine the sum of square of characteristic difference, and the number of local competitors. When using an unbiased first-stage parameter, the optimal IV approximation improves the precision of the estimates by 60% in the Hotelling model, and by 17% in the Normal RC model. However, these efficiency gains are quickly eliminated when the first-stage parameter is set far from

$\boldsymbol{\theta}^0$. This is an important advantage of the Differentiation IVs, since their exact structure does not depend on the availability of consistent estimates, or on prior the knowledge of the model of differentiation (e.g. Hotelling versus normal).

The previous examples are very stylized. Another setting in which the sign and magnitude of $\boldsymbol{\theta}$ determines substitution patterns is the correlated random-coefficient model studied in Section 3.4. To illustrate the importance of using consistent estimates in the first-stage, we implement the optimal IV approximation using pseudo-random values that are **not** centered around the truth. The point here is not to replicate the results from Reynaert and Verboven (2013), but rather to highlight the importance of using consistent estimates in the first stage.

The results are summarized in Table A2 in the Appendix. In columns (2)-(4), each element of $\boldsymbol{\theta}^1$ is drawn from a standard-normal distribution, while in columns (4)-(6) they are drawn from a normal distribution with a standard-deviation of 2. The results are in line with the single-address Hotelling example. Using inconsistent parameter estimates to approximate the optimal instruments leads to a weak identification problem, associated with very noisy and often biased parameter estimates. In addition, as we increase the variance of $\boldsymbol{\theta}^1$, the precision and bias of $\hat{\boldsymbol{\theta}}$ both increase substantially. The contrast with the Differentiation IVs is quite striking: the average RMSEs are roughly 5 times smaller with the Differentiation IVs than with the less noisy optimal IV approximation.

A valid strategy to improve the efficiency of the estimates is to obtain first-stage estimates using the instruments proposed in this paper, and then construct an approximation to the optimal IV. The second-stage can be conducted using the heuristic approximation discussed in Berry, Levinsohn, and Pakes (1999) and Reynaert and Verboven (2013), or using non-parametric regressions as discussed in Newey (1993) for instance.

We illustrate the performance of the former approach using the model with endogenous prices studied in Section 4.2. Table A3 in the Appendix summarizes the results. The top-panel corresponds to the GMM estimates obtained using three alternative Differentiation IV: (i) local competition, (ii) sum of square of characteristic differences, and (iii) sum of rival characteristics. In each specification we use the residual cost-shock, ω_{jt}, as a price instrument. In the bottom-panel, we use the GMM results from the corresponding specification to construct an approximation to the optimal IV, as described in equation (20). Each entry is averaged over 1,000 Monte-Carlo replications.

The results suggest that the Berry, Levinsohn, and Pakes (1999) approximation successfully corrects the weak identification problem. For instance, the sum of rival characteristics specification is associated with very noisy estimates of λ_p in the top panel, but the average bias and RMSE are mostly comparable across columns in the bottom panel. Similarly, the RMSE of λ_p estimated with the quadratic Differentiation IVs is roughly 50% smaller in the second-stage. The efficiency gains are much smaller in the first specification (17%), mostly because the local Differentiation IVs are

stronger instruments in this case.

Importantly, the simulation results illustrate a strong complementarity between the two approaches. The second-stage estimates are more precisely estimated when the Differentiation-IVs are used in the first-stage. In other words, using stronger instruments in the first stage lead to more precise results in the second stage. This should be thought of as a lower bound on the efficiency gains of using strong versus weak first-stage instruments. As we saw in the "Hotelling" vs "Normal RC" examples above (see Table 5) the efficiency loss from using inconsistent initial parameter values is small in the multiplicative random-coefficient specification. Also, Reynaert and Verboven (2013)'s simulation results suggest that the heuristic approximation is becoming weaker as the the number of random-coefficients increases beyond four. It is likely that the complementarity between the two approaches would increase with the number of non-linear parameters, since the performance of the Differentiation-IVs is very stable across different dimensions of consumer heterogeneity.

Conclusion

In this paper, we have analyzed the theoretical and small-sample properties of a new family of instruments used to estimate substitution patterns: the Differentiation IVs. We demonstrate that exogenous measures of differentiation (or proximity in characteristics) solves the weak identification problem associated with commonly used moment conditions.

Importantly, these instruments are derived from two common assumptions on the primitives of the demand model: (i) linear-in-characteristics indirect utility function, and (ii) exchangeability of the residual demand shocks. We use these two restrictions to establish that the reduced-form of the model is a vector-symmetric function of characteristic differences; a property that solves the curse-of-dimensionality problem in the reduced-form.

The results have important implications for applied work. The proposed instruments are low-dimension functions that are easy to construct, and have a close connections with theories of product differentiation. Our empirical simulations show that they are powerful even in moderately samples, and yield precise and unbiased of the key parameters determining the elasticity of substitution between products. We also show how the Differentiation IVs can be extended to identify models with endogenous characteristics, demographic characteristics, and correlated random-coefficients. These results are confirmed by a growing number of applications (including for instance: Miravete, Moral, and Thurk 2018, Coşar, Grieco, and Tintelnot 2018, Chaves 2019, and Singleton 2019), as well as recent monte-carlo simulation analysis (e.g. Conlon and Gortmaker (2019)).

Our approach to identification and estimation also suggest a natural methodology to conduct empirical work and report results when estimating demand for differentiated-products. Prior to estimating the model, researchers should first conduct an analysis of the reduced-form of the model, by estimating the IIA regression described in Section 1. Our simulation results demonstrate that the model is weakly identified if the IIA hypothesis cannot be rejected (only weakly rejected).

This test is easy to implement, and can be useful to help identifying strong moment restrictions. Furthermore, after estimating the model, Differentiation IVs can be used to conduct specification tests evaluating the validity of alternative modeling choices. In particular, the relevance of the instruments is independent of the assumptions regarding the distribution for the random-coefficients (e.g. normal, log-normal, correlated), or the functional form of the utility function (e.g. vertical form *vs* Hotelling). This feature allow researchers to conduct non-nested specification tests, based on the validity of the over-identification restrictions.

Finally, the results open several avenues for future research. For instance, the IIA regression can be used to identify more flexible basis functions capable of approximating the reduced-form. In many applications the number of product characteristics is large, and machine-learning methodologies can be used to select the "best" combinations of Differentiation IVs that can reject the IIA hypothesis.[28] Another potential application our symmetry result is for the estimation of semi-parametric models of demand for differentiated-products. As we have shown, symmetric models of demand do not suffer from a curse of dimensionality problem, and could in principle be estimated non-parametrically using an approach similar to Compiani (2019).

[28] See Gillen, Moon, Montero, and Shum (2019) for an analysis of LASSO methods in a related context.

A Proofs and additional derivations

A.1 IIA-regression and GMM

In this subsection we formally derive a minimum distance estimator based on the IIA-regression introduced in Section 1, and discuss its equivalence with GMM.

Let $A_j(\boldsymbol{x}_t)$ denotes a $1 \times L$ vector of basis functions summarizing the choice-set in market t. Importantly $A_j(\boldsymbol{x}_t)$ includes the vector of characteristics of product j ($\boldsymbol{x}_{jt}$).

The IIA-regression is obtained by projecting the inverse-demand under logit onto $A_j(\boldsymbol{x}_t)$:

$$E[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t] = \boldsymbol{x}_{jt}\boldsymbol{\beta}^0 + E\left[\Delta_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\theta^0\right)|\boldsymbol{x}_t\right] + \underbrace{E[\xi_{jt}|\boldsymbol{x}_t]}_{=0} = A_j(\boldsymbol{x}_t)\gamma_n \tag{22}$$

where $\Delta_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\theta}^0\right) = \sigma^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\lambda}=0\right) - \sigma^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\lambda}^0\right)$ is the deviation from IIA in the inverse-demand function, and $\boldsymbol{\theta}^0 = (\boldsymbol{\beta}^0, \boldsymbol{\lambda}^0)$ is the true vector of parameters.

The same conditional expectation evaluated at parameter $\boldsymbol{\theta}$ can be estimated from the following regression:

$$E\left[y_{jt}(\boldsymbol{\theta})|\boldsymbol{x}_t\right] \approx A_j(\boldsymbol{x}_t)\boldsymbol{\gamma}(\boldsymbol{\theta}) \tag{23}$$

where $y_{jt}(\boldsymbol{\theta}) = \sigma^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\lambda}=0\right) - \sigma^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}|\boldsymbol{\lambda}\right) + \boldsymbol{x}_{jt}\boldsymbol{\beta}$, and $\boldsymbol{\gamma}(\boldsymbol{\theta})$ is the regression coefficient vector obtained by regressing $y_{jt}(\boldsymbol{\theta})$ on $A_j(\boldsymbol{x}_t)$.

A minimum-distance estimator can be constructed by matching the IIA reduced-form model estimated from the data, with the model-based prediction obtained from equation (23). This is a valid estimator of the parameters since the two conditional expectations are equal when evaluated at the true parameter value:

$$\begin{aligned} & E\left[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t\right] - E\left[y_{jt}(\boldsymbol{\theta})|\boldsymbol{x}_t\right] = 0 \quad \text{If } \boldsymbol{\theta} = \boldsymbol{\theta}^0 \\ \rightarrow\ & E\left[\underbrace{A_{jt}(\boldsymbol{x}_t)\boldsymbol{\gamma}_n - A_{jt}(\boldsymbol{x}_t)\boldsymbol{\gamma}(\boldsymbol{\theta}^0)}_{e_{jt}(\boldsymbol{\theta}^0)}\right] = 0. \end{aligned}$$

The minimum-distance estimator of $\boldsymbol{\theta}$ is obtained by minimizing the (weighted) sum of square of residuals $e_{jt}(\boldsymbol{\theta})$:

$$\boldsymbol{\theta}^{MD} = \min_{\theta} \quad \boldsymbol{e}(\boldsymbol{\theta})'\Sigma^{-1}\boldsymbol{e}(\boldsymbol{\theta})$$

In contrast, the GMM estimator proposed by Berry et al. (1995) is given by:

$$\min_{\boldsymbol{\theta}} \quad g_n(\boldsymbol{\theta})W g_n(\boldsymbol{\theta})^T \tag{24}$$

where $g_n(\boldsymbol{\theta}) = \rho(\boldsymbol{s}, \boldsymbol{x}|\boldsymbol{\theta})^T A(\boldsymbol{x}) = \left(\sigma^{-1}(\boldsymbol{s}, \boldsymbol{x}^2|\boldsymbol{\lambda}) - \boldsymbol{x}\boldsymbol{\beta}\right)^T A(\boldsymbol{x})$.

Recall that the two conditional expectations in equations (22) and (23) are obtained using the same projection matrix: $P_A = A(A^TA)^{-1}A^T$. The residual from the minimum distance estimator can thus be re-written as:

$$e_{jt} = P_A \left[\ln s_{jt}/s_{0t} - \left(\ln s_{jt}/s_{0t} - \sigma^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)} | \boldsymbol{\lambda}\right) + \boldsymbol{x}_{jt}\boldsymbol{\beta} \right) \right] = P_A \rho(\boldsymbol{\theta})$$

Therefore, the minimum-distance estimator is equivalent to a GMM problem with a "sandwich" weighting matrix given by: $W = (A^TA)^{-1}A^T\Sigma^{-1}A(A^TA)^{-1}$.

A.2 First Proposition

Proposition 1 can be restated as follows. For simplicity we remove the t subscript associated with each matrix, and drop the parameter vector from the conditioning variables.

Proposition 1. *In the linear characteristics model the market inverse function can be expressed as*

$$\sigma_j^{-1}\left(s_0, s_1, \ldots, s_J; \boldsymbol{x}\right) \quad = \quad G\left(s_j, \left\{s_k, \boldsymbol{d}_{jk}^{(2)}\right\}_{k \neq j}\right) + C$$

where $\boldsymbol{d}_{jk} = \boldsymbol{x}_k - \boldsymbol{x}_j$ *and* C *is a constant that is common to all products* $j = 1, \ldots, J$.

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

$$G\left(s_j, \{s_k, \boldsymbol{d}_{jk}\}_{k \neq j}\right) = G\left(s_j, \boldsymbol{F}_j\left(s, d^{(2)}\right)\right)$$

where we have equivalently expressed the second argument as the empirical distribution of $\left(s_k, \boldsymbol{d}_{jk}^{(2)}\right)$ among products $k \neq j$ (which includes the outside good 0 in this sample). It is important to note that from this empirical distribution, we can only recover the set of the differences d_{jk} but cannot isolate the difference with respect to any particular product, and also cannot recover x_j itself from this distribution (because we cannot identify the outside good in this set). This brings to light that the cross sectional variation in the inverse function does not actually depend on a product's level of own x_j, but rather the distribution of differences d_{jk} for $k \neq j$ this product faces.

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

Step 1

The first step is to re-parameterize the demand function $\sigma_j\left(\delta_1, \ldots, \delta_J, \boldsymbol{x}^{(2)}\right)$ in terms of

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

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

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

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

Let $\boldsymbol{\theta}_i = (v_{i1}, \ldots, v_{iK_2}, \epsilon_{i0}, \ldots, \epsilon_{iJ})$ denotes the vector consumer taste parameters with joint CDF $\Phi(\cdot)$. We can thus express demand in terms of this re-parameterization, i.e.,

$$u(t_j, \boldsymbol{x}_j, \boldsymbol{\theta}_i) = T_j + \sum_{k=1}^{K_2} v_{ik} x_{jk}^{(2)} + \epsilon_{ij}$$

and

$$\mathcal{D}_j(t_0, \ldots, t_J) = \int \mathbf{1}\left[u(t_j, x_j, \theta) \geq u(t_k, x_k, \theta) \ \ \forall k = 0, \ldots, J, \ \ k \neq j\right] \ d\Phi(\theta). \tag{25}$$

We then have that

$$\mathcal{D}_j\left(t_0, \ldots, t_J, \boldsymbol{x}^{(2)}\right) = D_j\left(\delta_1, \ldots, \delta_J, \boldsymbol{x}^{(2)}\right).$$

This is because preferences are translation invariant. Moreover we have that

$$\log \mathcal{D}_j^{-1}\left(s_0, \ldots, s_J, \boldsymbol{x}^{(2)}\right) + C = D_j^{-1}\left(s_0, \ldots, s_J, \boldsymbol{x}^{(2)}\right).$$

Our strategy moving forward is to show that

$$\mathcal{D}_j^{-1}\left(s_0, \ldots, s_J, \boldsymbol{x}^{(2)}\right) = \mathcal{D}^{-1}\left(s_j, \left\{s_k, \boldsymbol{d}_{jk}^{(2)}\right\}_{k \neq j}\right). \tag{26}$$

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

Step 2

We now establish 3 properties of $\mathcal{D}_j\left(t_1, \ldots, t_J, \boldsymbol{x}^{(2)}\right)$: symmetry, anonymity, and translation invariance. Each of these properties will then be preserved by the inverse mapping $\mathcal{D}_j^{-1}$. To establish these properties let us define a product j's state $\boldsymbol{\omega}_j$ as

$$\boldsymbol{\omega}_j = (t_j, \boldsymbol{x}_j)$$

and note that

$$\mathcal{D}_j\left(t_0, \ldots, t_J, \boldsymbol{x}^{(2)}\right) = \mathcal{D}_j\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right).$$

The following two properties are relatively straightforward to show using the definition of demand (25) and the symmetry of the idiosyncratic errors (ϵ_{ij}). The first property is

Definition 3. *The function* $\mathcal{D}_j\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$ *is* symmetric *if* $\mathcal{D}_j\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right) = \mathcal{D}_k\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$ *for any* $k \neq j$.

This implies we can write $\mathcal{D}_j\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right) = \mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$.

Definition 4. *The function* $\mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$ *is anonymous if* $\mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{\rho(-j)}, \boldsymbol{x}^{(2)}\right)$ *where* ρ *is any permutation of the indices* $-j$.

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

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

Definition 5. *The function* $\mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$ *is translation invariant if for any* $c \in \mathbb{R}^K$ *we have that*

$$\mathcal{D}\left(\boldsymbol{\omega}_j + (0, c), \boldsymbol{\omega}_{-j} + (\vec{0, c}), \boldsymbol{x}^{(2)}\right) = \mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$$

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

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

Step 3

Now define the relevant state for the inverse mapping as

$$\boldsymbol{m}_j = (s_j, \boldsymbol{x}_j).$$

Then

$$\mathcal{D}_j^{-1}\left(s_0, \ldots, s_J, \boldsymbol{x}^{(2)}\right) = \mathcal{D}_j^{-1}\left(m_j, m_{-j}\right).$$

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

$$\begin{aligned}
\mathcal{D}_j^{-1}\left(\boldsymbol{m}_j, \boldsymbol{m}_{-j}\right) &= \mathcal{D}^{-1}\left(\boldsymbol{m}_j + \left\langle 0, -\boldsymbol{x}_j^{(2)}\right\rangle, \boldsymbol{m}_{-j} + \left\langle 0, -\boldsymbol{x}_j^{(2)}\right\rangle\right) \\
&= \mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k \neq j}\right)
\end{aligned}$$

where the first equality follows from symmetry and translation invariance, and the second equality follows from anonymity. We have thus succeeded in establishing (26) and hence Theorem 1.

A.3 Second Proposition

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

Proposition 2. *The conditional expectation of interest in the model can be expressed as*

$$\begin{aligned} E\left[\mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k\neq j}\right) \mid \boldsymbol{x}\right] &= E\left[\mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k\neq j}\right) \mid \{\boldsymbol{d}_{jk}\}_{k\neq j}\right] \\ &= E\left[\mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k\neq j}\right) \mid F_j\left(d\right)\right] \end{aligned}$$

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

Assume that the d_{jk} can be canonically ordered (based on some complete ordering in $\mathbb{R}^K$, such as the lexicographic ordering) such that $\tilde{\boldsymbol{d}}_{j1} \leq \cdots \leq \tilde{\boldsymbol{d}}_{jK}$ where $\tilde{\boldsymbol{d}}_{jl}$ is the lth largest from the $\{\boldsymbol{d}_{jk}\}_{k\neq k}$. Then we can express

$$\mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k\neq j}\right) = D^{-1}\left(\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ}; \tilde{\boldsymbol{d}}_{j1}^{(2)}, \ldots, \tilde{\boldsymbol{d}}_{jJ}^{(2)}\right)$$

where $\tilde{s}_{j0}$ is s_j and $\tilde{s}_{ji}$ is the market share corresponding to the product with difference $\tilde{\boldsymbol{d}}_{ji}^{(2)}$. Now it can be shown that the distribution

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

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

$$\begin{aligned} E\left[\mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k\neq j}\right) \mid \boldsymbol{x}\right] &= E\left[D^{-1}\left(\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ}; \tilde{\boldsymbol{d}}_{j1}^{(2)}, \ldots, \tilde{\boldsymbol{d}}_{jJ}^{(2)}\right) \mid \boldsymbol{x}\right] \\ &= E\left[D^{-1}\left(\tilde{s}_{j0}, \tilde{s}_{j1}, \ldots, \tilde{s}_{jJ}; \tilde{\boldsymbol{d}}_{j1}^{(2)}, \ldots, \tilde{\boldsymbol{d}}_{jJ}^{(2)}\right) \mid \tilde{\boldsymbol{d}}_{j1}, \ldots, \tilde{\boldsymbol{d}}_{jJ}\right] \\ &= E\left[\mathcal{D}^{-1}\left(s_j, \left\{\left(s_j, \boldsymbol{d}_{jk}^{(2)}\right)\right\}_{k\neq j}\right) \mid F_j\left(\boldsymbol{d}\right)\right] \end{aligned}$$

A.4 Derivation of example with demographic differences

Consider the following single dimension example (Nevo 2001):

$$u_{ijt} = \delta_{jt} + b_{it} x_{jt}^{(2)} + \varepsilon_{ijt}, \quad b_{it} = \lambda_y y_{it} + \eta_i. \tag{27}$$

The random coefficient is composed of a demographic component y_{it} that is distributed according to (known) CDF $D_t(y)$, and a residual component ν_i that is normally distributed with mean zero and variance λ_η^2. The vector of non-linear parameters contains two elements: $\boldsymbol{\lambda} = \{\lambda_y, \lambda_\eta\}$.

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

$$y_{it} = m_t + \text{sd}_t e_i \text{ such that } \Pr(e_i < x) = \Psi_e(x).$$

where $\{m_t, \text{sd}_t\}_{t=1,\ldots,T}$ and $\Psi_e(x)$ are known transformation of the observed distribution $D_t(y)$.

We can use this standardization to express the aggregate demand function:

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

where $\tilde{\boldsymbol{x}}_{jt}^{(2)} = \left\{\text{sd}_t x_{jt}^{(2)}, x_{jt}^{(2)}\right\}$ is an expanded vector of non-linear characteristics, $\boldsymbol{v}_i = \{e_i, \eta_i\}$, and is the joint density of $\boldsymbol{v}_i$ defined from $\phi_\eta(\cdot)$ and $\Psi_e(\cdot)$.

Note that the change of variables allows us to eliminate the t subscript from the demand function, and expand the state space by adding two new interactions: (i) the mean of y_{it} times $x_{jt}^{(2)}$, and (ii) the standard-deviation of y_{it} times $x_{jt}^{(2)}$.

Under this new parametrization of the model, we can use directly Proposition 2 to write the reduced-form of the model as follows:

$$\pi_{jt}(\boldsymbol{x}_t; \boldsymbol{\lambda}) = g\left(\boldsymbol{d}_{jt}, \text{sd}_t \cdot \boldsymbol{d}_{jt}^{(2)}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda}) \tag{28}$$

B Monte Carlo Simulation Designs and Algorithms

B.1 Monte Carlo Simulations

We use the following parametrization for the independent random-coefficients specifications:

- $\boldsymbol{\Sigma}_{k,k} = \lambda = 4$ for all $k = 1, \ldots, K_2$
- $\nu_{ik} \sim N(0,1)$ for all $k = 1, \ldots, K_2$
- $\epsilon_{ij} \sim \text{T1EV}(0,1)$

We use the following covariance matrix for the correlated random-coefficient example:

Table 6: Random Coefficient Covariance Matrix

	c1	c2	c3	c4
r1	4.000			
r2	-2.000	4.000		
r3	2.000	-2.000	4.000	
r4	2.000	-2.000	2.000	4.000

The data-generating process for all numerical exogenous characteristics examples in Sections 3.4 and 4.2 is described as:

- Number of products (J_t): 15
- Number of market (T): 100
- Observed characteristics: $x_{jt,k} \sim N(0,1)$ for all $k = 1, \ldots, K$
- Cost shifter: $\omega_{jt} \sim N(0,1)$
- Unobserved quality: $\xi_{jt} \sim N(0,1)$

B.2 Computational Procedure

All numerical simulations and optimizations were done using the matrix programming language Ox (Doornik 2007). We use a nested fixed-point algorithm to solve the non-linear GMM problem:

$$\min_{\boldsymbol{\theta}} \quad n g_n(\boldsymbol{\theta}) \boldsymbol{W}_n g_n(\boldsymbol{\theta})^T \tag{29}$$

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

The residual function is obtained by inverting the demand function for a candidate parameter vector $\boldsymbol{\lambda}$. We use the following Newton-Raphson method root-finding algorithm to solve this problem separately for each market t. Following Berry et al. 1995, the algorithm solves the following non-linear system of equation:

$$f_j(\boldsymbol{\delta}) = \ln s_{jt} - \ln \sigma_j(\boldsymbol{\delta}_t, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda}) = 0 \quad \forall j = 1, \ldots, J. \tag{30}$$

Algorithm 1 (Demand Inversion)**.** *Initiate the algorithm at vector of quality $\boldsymbol{\delta}_t^1$ (e.g. solution evaluated at last iteration parameter's guess). Iteration l:*

1. *Evaluate the predicted demand via Monte-Carlo simulation:*

$$\sigma_j(\boldsymbol{\delta}_t^k, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda}) = \frac{1}{S} \sum_i \frac{\exp(\delta_{jt}^l + \sum_k \boldsymbol{\lambda}_k \nu_{ik} x_{ij,k})}{1 + \sum_{j'} \exp(\delta_{j't}^l + \sum_k \boldsymbol{\lambda}_k \nu_{ik} x_{ij',k})}$$

2. *Use the implicit theorem to calculate the* $J \times |\boldsymbol{\lambda}|$ *Jacobian matrix of the zero-function* $f(\boldsymbol{\delta})$ *above:*

$$F(\boldsymbol{\delta}^l) = \left\langle -1/\boldsymbol{\sigma}(\boldsymbol{\delta}^l, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda}) \right\rangle \circ \left\langle -\frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta}^l, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda})}{\partial \boldsymbol{\delta}^T} \right\rangle^{-1} \left\langle \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\delta}^l, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right\rangle$$

3. *Updating:*

$$\boldsymbol{\delta}^{l+1} = \begin{cases} \boldsymbol{\delta}^l + f(\boldsymbol{\delta}^l) & \text{If } ||f(\boldsymbol{\delta}^l)|| > \epsilon^1 \\ \boldsymbol{\delta}^l + F(\boldsymbol{\delta}^l)^{-1} f(\boldsymbol{\delta}^l) & \text{If } ||f(\boldsymbol{\delta}^l)|| \leq \epsilon^1 \end{cases}$$

4. *If* $||f(\boldsymbol{\delta}^l)|| < \epsilon^2$*, stop. Else repeat step 1-3.*

This root-finding algorithm use two tolerance variables (ϵ^1 and ϵ^2). The first one determines the threshold after which the algorithm starts to use Newton-Raphson steps. We set $\epsilon^1 = 0.1$. When this value is increased, the algorithm is equivalent to the contraction-mapping algorithm proposed by Berry et al. 1995. The advantage of the Newton-Raphson steps is that it converges at a faster rate than the contraction-mapping. However, it can diverge when the starting values are too far from the truth. We set the overall convergence criteria equal to: $\epsilon^2 = 10^{-16}$. Note also that this algorithm is easily parallelizable, since a fixed-point vector needs to be calculate separately for each market.

Since the GMM objective function is a quadratic form, the Gauss-Newton Regression (GNR) algorithm is a computationally efficient method for finding the minimum (see for instance Newey (1993)). Each optimization step is obtained by estimating a linear GMM problem corresponding to a linear approximation of the residual function.

Algorithm 2 (Gauss-Newton Regression)**.** *Initiate the algorithm at parameter* $\boldsymbol{\theta}^1$*. Iteration* k*:*

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

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

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

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

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

4. *Update parameter vector:*

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

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

The Gauss-Newton algorithm has good convergence properties when the moments are strong. This is because strong instruments imply a lot of curvature in the GMM objective function, which is therefore well approximated by a quadratic function. In contrast, weak instruments are associated with little or no curvature in the objective function, which leads to convergence problems. We use the GNR algorithm in all specifications using Differentiation IVs. To estimate the model with weak instruments, we use a Nelder-Mead (or Simplex) algorithm to find the local minimum

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

$$\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}) = \boldsymbol{Z}\boldsymbol{\pi}\boldsymbol{b} + \boldsymbol{v}^1 \tag{31}$$

$$\boldsymbol{J}(\hat{\boldsymbol{\theta}}) = \boldsymbol{Z}\boldsymbol{\pi} + \boldsymbol{v}^2 \tag{32}$$

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

C Additional Tables

Table A1: Monte-Carlo simulation results for exogenous characteristics model with strong instruments

(a) Differentiation IV: Quadratic

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

(b) Differentiation IV: Local

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

Table A2: Monte-Carlo simulation results for correlated random-coefficient specification with optimal IV approximation and inconsistent initial parameter values

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

Table A3: Monte-Carlo simulation results for endogenous price specification and optimal IV approximation

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

References

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

Altonji, Joseph, G. and L. Matzkin, Rosa (2005, July). Cross section and panel data estimators for nonseparable modesl with endogenous regressors. *Econometrica 73*(4), 1053–1102.

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

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

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

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

Armstrong, T. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply". *Econometrica 84*(5), 1961–1980.

Bayer, P., F. Ferreira, and R. McMillan (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy 115*(5), 588–638.

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, Feb). Differentiated products demand systems from a combination of micro and macro data: The new car market. *The Journal of Political Economy 112*(1), 68.

Berry, S. and A. Pakes (2007). The pure characteristics demand model. *International Economic Review 48*(4), 1193–1225.

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.

Berry, S. T. and P. Haile (2016). Identification in differentiated products markets. *Annnual Review of Economics 8*, 27–52.

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

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

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

Chaves, D. (2019, January). Taxation and product variety: Evidence fromthe brazilian automobile industry. Working paper, University of Wisconsin-Madison.

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

Coşar, K. A., P. L. Grieco, and F. Tintelnot (2018). What drives home market advantage. *Journal of International Economics 110*, 135–150.

Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: A nonparametric approach. working paper, Berkeley University.

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

Conlon, C. and J. Gortmaker (2019, May). Best practices for differentiated products demand estimation with pyblp. working paper, NYU-Stern.

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

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

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

Doornik, J. A. (2007). *Ox - An Object-Oriented Matrix Programming Language*. Timberlake Consultants.

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

Dube, J.-P., J. Fox, and C.-L. Su (2012, September). Improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation. *Econometrica*.

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

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

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

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

Gillen, B., R. Moon, S. Montero, and M. Shum (2019). "blp-2lasso" for aggregate discrete-choice models with rich covariates. *Econometrics Journal*.

Hausman, J. A. (1994). Valuation of new goods under perfect and imperfect competition. NBER wp 4970.

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

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

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

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

Miravete, E., M. Moral, and J. Thurk (2018, Fall). Fuel taxation, emissions policy, and competitive advantage in the diffusion of european diesel automobiles. *Rand Journal of Economics*, 504–540.

Miravete, E., K. Seim, and J. Thurk (2018). Market power and the laffer curve. *Econometrica 86*(5), 1651–1687.

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

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

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

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Nielson, C. (2017). Targeted vouchers, competition among schools, and the academic achievement of poor students. Working paper, Princeton University.

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

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

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

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

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

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

Singleton, J. D. (2019). Incentives and the supply of effective charter schools. *American Economic Review 109*(7), 2568–2612.

Starc, A. (2014). Insurer pricing and consumer welfare: evidence from medigap. *RAND Journal of Economics 45*(1), 198–220.

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

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

Stock, James H, Y. M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pp. 80–108. Identification and Inference for Econometric Models.

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

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

Waldfogel, J. (2003, Autumn). Preference externalities: An empirical study of who benefits whom in differentiated-product markets. *RAND Journal of Economics 34*(3), 557–568.

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