MEASURING SUBSTITUTION PATTERNS IN DIFFERENTIATED-PRODUCTS INDUSTRIES

Amit Gandhi
Jean-François Houde

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 demonstrate the consequences of using weak instruments in this non-linear GMM context, and propose a new class of instruments that are designed to avoid weak IV and can be used to estimate a large family of models with aggregate data. We argue that strong 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.

Amit Gandhi
University of Pennsylvania
akgandhi@sas.upenn.edu

Jean-François Houde
Department of Economics
University of Wisconsin-Madison
1180 Observatory Dr
Madison, WI 53706
and NBER
houde@wisc.edu

Introduction

The extent to which competing products are substitutable is central to empirical Industrial Organization (IO) because it is informative about 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 however complicates the identification and estimation of substitution patterns, since the introduction of non-IIA preferences creates a simultaneity problem associated with the joint determination of market shares and unobserved attributes. For instance, in a Nested-Logit model (special case of the random-coefficient model), the price and market shares of products in the same nest are correlated with the model residual; hence the need for separate Instrumental Variables (IV) (Berry 1994). In the random-coefficient model, market-shares enter the model non-linearly, and the parameters governing substitution patterns are estimated using non-linear IV estimators. This class of estimators is notoriously sensitive to the presence of weak identification; a problem that is difficult to diagnose (e.g. Stock and Wright 2000).

The problem of weak identification has received little attention in the demand estimation literature, and a review of the empirical literature suggests that it is a 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 IIA preferences relying solely on demand restrictions.[2] In addition, commonly used moment conditions often lead to numerical optimization problems; another symptom of weak identification in non-linear models.[3]

---

[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]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).

[3]See Metaxoglou and Knittel (2014) and Dube, Fox, and Su (2012) for a discussion of numerical

Our goal in this paper is to develop a new class of *strong* instruments for estimating substitution patterns of demand. Strong instruments are relevant instruments that can flexibly estimate substitution patterns while also circumventing the weak identification challenges above.[4]

In order to build strong instruments, we first theoretically examine the source of weak identification. Weak identification in non-linear problems arises when the residual function is weakly correlated with the instruments away from the true parameter value. As with linear models, the strength of this correlation is measured by regressing the residual function (endogenous variables) on the instruments. We refer to this regression as the "reduced form" of the model.

The main challenge for empirical work is that the strength of the reduced form relationship cannot be evaluated empirically without knowing the values of the parameters, e.g., there is no clear analogue of a first-stage regression that is typically employed for linear models in the non-linear context. In the demand estimation setting, this is compounded by the fact that the number of exogenous variables grows with number of products in a market, and therefore with the sample size.

Our main theoretical contribution is to show that this 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 and exchangeable errors.

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). This implies that relevant instruments should measure the degree of differentiation of a product relative to others available in the market. Importantly the *number* of basis functions necessary to explain the (unknown) reduced-form is invariant to the number of products in the market. We propose a series of instrument functions that satisfy this property, and label them

---

problems/solutions in this context.

[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.

Differentiation IVs.

Our second contribution is to the detection of weak instruments. We show that an instrument is weak if it fails to reject the null hypothesis of IIA preferences. This hypothesis can be tested by estimating the (linear) reduced-form of the model under Logit preferences. This reduced-form is also a symmetric function of the distribution of characteristics differences, and Differentiation IVs can therefore be used to test the IIA hypothesis. Importantly this test is not accompanied by a statistical criteria to determine when an instrument function is "too weak". However, it provides a useful approach the analyze empirically the sign and strength of the reduced-form relationships; despite the presence of a large number of endogenous variables, and the non-linearity of the model.

These two results suggest a sequential approach estimation. First, the theory can be used to construct candidate instruments; the form of which can vary depending the structure of the data. Second, researchers can evaluate the strength of the proposed instruments by testing the IIA hypothesis. We illustrate this approach empirically using Monte-Carlo simulations, and by estimating a model of demand for new cars.

Our simulation results show that the proposed instruments, by eliminating the weak IV problem, can improve substantially the precision of the estimates (by a factor of 10 in some cases), and the numerical performance and speed of the non-linear optimization algorithms used to estimate the parameters. We discuss how to adapt our identification strategies to settings with correlated random-coefficients, endogenous product attributes and large dimension problems. We also compare our results with the Optimal IV approximation proposed by Berry, Levinsohn, and Pakes (1999), and a model with endogenous attributes identified using quasi-experiment variation.

We then illustrate the IIA test by revisiting the car application first studied by Berry et al. (1995). We find statistically significant deviations from IIA along (at least) three dimensions: price, car size (doors), and air-conditioning. Using measures of differentiation along these three dimensions as instruments, and show that it is possible to precisely estimate the random-coefficient parameters (controlling for brand and market fixed-effects), without relying on restrictions external to the demand model (e.g. supply-side or external survey data). This is important, since ability to identify substitution patterns without relying on an equilibrium model of supply is central to the analysis of firm conduct in Industrial Organization (Bresnahan 1982, Berry and Haile 2014).

3

A growing number of papers using our instruments confirm these results. The Differentiation IVs introduced in this paper have now been used to obtain precise estimates of substitution patterns in a variety of applied contexts, ranging from demand for cars (Miravete, Moral, and Thurk 2018, and Coşar, Grieco, and Tintelnot 2018), scanner data (Miller and Weinberg (2017), Sullivan (2020), Dube, Hortacsu, and Joo (2020)), and school choice (Singleton 2019). See also Conlon and Gortmaker (2019) for additional Monte-Carlo simulation results.

Our paper is related to a set of recent papers that have also raised concerns about the efficiency of instruments used in standard practice and explored alternative instrumental variable strategies for differentiated product demand models. Many of these papers are focused on approximating the *optimal* instruments for GMM under the conditional moment restrictions of the model in the sense of Amemiya (1977) and Chamberlain (1987). For example, Reynaert and Verboven (2013) discussed the loss of efficiency associated with commonly used instruments, and analyze the small sample performance of the approximation to the optimal instruments proposed by Berry et al. (1999). Salanie and Wolak (2019) propose an estimation and specification selection procedure based on the Gauss-Newton regression, which also rely on the Jacobian of the residual function.

The challenge with these approaches is that they require a starting value for the non-linear parameters that are in a reasonable vicinity of the true value, which itself requires an initial choice of instruments by the researcher to determine an initial value. If the initial instrumental variable are weak, these optimal IV refinements will have a difficult time escaping the weak-IV regime.

Our approach is ultimately complementary to these methodologies by focusing on the weak instruments problem directly. Differentiation IVs can be used as in input to these procedures by obtaining initial estimates that are not subject to the weak IV problems. An approximation to the Optimal IV using these approaches can be used to obtain further efficiency gains. As discussed in Conlon and Gortmaker (2019), this is now considered the empirical "best practice" for estimating this class of models. However Differentiation IV's have applications beyond sheer efficiency of estimation and can be used to estimate, test, and compare different models of product differentiation. This arises because a key advantage of our approach is that our characterization of the reduced-form *does not* depend on the distribution of the random-coefficient, or on the value of the parameters. Having instruments that are agnostic to the parametrization

of demand heterogeneity allows researchers to to test alternative specifications and gain applied insights into the appropriate structure of demand for the data at hand.

The instruments that we propose are also similar to the instruments commonly used to identify nested-logit and spatial differentiation models.[5] A key contribution of our paper is to show that the intuition underlying these instruments for the particular differentiation structure they apply generalizes to the broader family family of characteristics models with random coefficients. We also derive an instrument function that can exploit variation in demographic characteristics across markets, similar to the one proposed by Romeo (2010).

The rest of the paper proceeds as follows. In the next section, we describe the model, and formally define the weak-identification problem. Sections 2 and 3 present our main theory results. We first derive our main theoretical result, and illustrate its implication for parametric and non-parametric estimation of the model. We then present a series of Monte-Carlo simulations in section 4 to analyze the finite-sample properties of the instruments. Section 5 applies the instruments to the estimation of a mixed-logit model of demand for new cars. The appendix include the proof of the main propositions, and computation details related to the Monte-Carlo simulations and application.

1 Description of the model

In this section we introduce the general notation that we will use throughout the paper. We also describe the identifying assumptions and the GMM estimator.

1.1 Model assumptions

Consider a panel data-set summarizing the characteristics and demand of differentiated products in T independent markets. Each market t is composed of J_t products, and 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). The vector of prices for market t are denoted by $\boldsymbol{p}_t = \{p_{1t}, \ldots, p_{J_t}\}$. Similarly, $\boldsymbol{s}_t =$

---

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

$\{s_{1t}, \ldots, s_{J_t}\}$ 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 $\boldsymbol{x}_{0t} = 0$.[6] We impose three assumptions on the model and data-generating process.

Our first assumption, refers to the shape of the indirect utility function generating choices. 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 1. *Each consumer i has linear preferences for products $j = 0, 1, \ldots, J_t$:*

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

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), ϵ_{ijt} is an IID random-utility shock for product j, and $\boldsymbol{\nu}_i = \{\nu_{ip}, \nu_{i2}, \ldots, \nu_{iK_2}\}$ is the vector of random-coefficients for consumer i.

When ϵ_{ijt} is distributed according to a T1EV distribution, the aggregate demand function for product j can be written as follows:

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

where $F(\boldsymbol{\nu}_i; \boldsymbol{\lambda})$ denotes the joint distribution of the random-coefficient vector in market t, where $\boldsymbol{x}_t^{(2)} = \left(\boldsymbol{x}_{1t}^{(2)}, \ldots, \boldsymbol{x}_{J_t,t}^{(2)}\right)$ and $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{J_t,t})$. 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 the pure-characteristic and semi-parametric demand models with linear preferences (e.g. Berry and Pakes 2007 and Compiani 2019).

Our second assumption is related to the conditional distribution of the model residual. Following Berry et al. (1995) and Berry and Haile (2014), we assume that ξ_{jt} vary independently of the choice-set that consumers face, and vector of price instruments $\boldsymbol{w}_t$. Assumption 2 formalizes this identifying restriction.

---

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

Assumption 2. *The unobserved quality of products has mean zero conditional on the observed menu of characteristics $\boldsymbol{x}_t$ and price instruments $\boldsymbol{w}_t$,*

$$E\left[\xi_{jt} \mid \boldsymbol{x}_t, \boldsymbol{w}_t\right] = 0. \tag{3}$$

Two broad sources of variation have been proposed in the literature to construct valid price instruments: (i) markup-shifters, and (ii) cost-shifters. See Berry and Haile (2016) for a review. In this paper we assume the researcher has access to excluded price instruments, and focus instead on constructing instrumental variables for market-shares.

Finally, our third assumption further restricts the joint distribution of the error term $\{\xi_{1t}, \ldots, \xi_{J_t,t}\}$.

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

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

for any ordering function $\iota()$.

In economics terms, this assumption implies that the identity of rival products is not important to predict the distribution of unobservable attributes. This assumption is not novel in the literature, and allow us to pool moment conditions across products. See discussion in Berry et al. 1995 (section 5.1). This assumption does not rule out the possibility that brand or product fixed-effects are relevant for consumers' decisions, and enter the indirect utility function linearly (e.g. Nevo (2001)).

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

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

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$ is the full parameter vector of dimension m. The inverse-demand function assigns quality levels to each product in order to matches the observed market-shares, for a given guess of the parameter $\boldsymbol{\lambda}$. Existence and uniqueness of this inverse

mapping follows directly from Berry et al. (1995).[7] A simultaneity problem arises because market shares (and prices) of rival products enter the inverse-demand non-linearly, and are correlated with the model residual. Importantly, even with exogenous prices, non-linear least square leads to a biased estimator of $(\boldsymbol{\lambda}, \boldsymbol{\beta})$.[8]

Theorem 1 in Berry and Haile (2014), shows that the parameters can be identified by imposing the following conditional-moment restrictions (CMR):

$$
\begin{aligned}
r_{jt}(\boldsymbol{\theta}) &= E\left[\rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}) \big| \boldsymbol{x}_t, \boldsymbol{w}_t\right] \\
&= E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}\right) \big| \boldsymbol{x}_t, \boldsymbol{w}_t\right] - \boldsymbol{x}_t \boldsymbol{\beta} = 0 \quad \text{iff } \boldsymbol{\theta} = \boldsymbol{\theta}^0.
\end{aligned}
\tag{5}
$$

The second term calculates the expectation of product j's inverse-demand, conditional 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 and prices.

1.2 Instrument relevance and the curse of dimensionality

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

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

where $\boldsymbol{z}_{jt} = \{\boldsymbol{x}_{jt}, A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)\}$. The *instrument function*, $A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)$, defines the set of excluded instruments characterizing the menu of characteristics in market t.

How should the instrument function be chosen? In linear models, excluded instruments are *relevant* if they exhibit a strong reduced-form correlation with the endogenous variables of the model. In non-linear models, the role of the instruments is to

---

[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, consider a model without prices. In this case, 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 $\sigma_j^{-1}(\cdot)$:

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

This echoes the discussions in Jorgensen and Laffont (1974) and Amemiya (1974).

approximate $r_j(\boldsymbol{\theta})$ from equation (5):

$$r_j(\boldsymbol{\theta}) \approx L\left[\rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta})\big|\boldsymbol{x}_t, \boldsymbol{w}_t\right] = l(\boldsymbol{z}_{jt}; \boldsymbol{\theta}),$$

where $l(\boldsymbol{z}_{jt}; \boldsymbol{\theta})$ is a projection of $\rho_{jt}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\theta})$ onto $\boldsymbol{z}_{jt}$. The GMM identification condition can be restated in terms of a zero function:

$$l(\boldsymbol{z}_{jt}; \boldsymbol{\theta}) = 0 \quad \text{if} \quad \boldsymbol{\theta} = \boldsymbol{\theta}^0.$$

Weak identification arises when this condition is nearly satisfied in the population away from true parameter value:

$$l(\boldsymbol{z}_{jt}; \boldsymbol{\theta}) \approx 0 \quad \text{if} \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}^0.$$

This leads to an empirical problem: In finite sample it is difficult to detect departures from $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ and the confidence intervals become large (Stock and Wright 2000).

The avoid this problem, researchers need to find a vector of excluded instruments, $A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)$, that can explain the residual function at an arbitrary parameter $\boldsymbol{\theta}$. The upper bound on the strength of the instruments is given by $r_j(\boldsymbol{x}_t, \boldsymbol{w}_t; \boldsymbol{\lambda})$. Ideally, the chosen instrument function can approximate the function arbitrarily well. Formally, Donald, Imbens, and Newey (2008) show that an instrument function is efficient if it satisfies a "spanning condition": it can approximate any function $a_j(\boldsymbol{x}_t, \boldsymbol{w}_t)$ arbitrarily well as the dimensionality of $A_j(\cdot)$ goes to infinity.

As discussed in Donald et al. (2008), "low-order approximating functions (e.g. linear or quadratic) often provide the most information". Without imposing further restrictions, this is not the case in the differentiated product context, since even low dimension basis functions suffer from a curse of dimensionality problem. 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. Recall that the number of exogenous variables $(\boldsymbol{x}_t, \boldsymbol{w}_t)$ is equal to $K \times J_t$, and the number of endogenous variables in the structural equation is equal to $2 \times J_t$. In many applications the number of products is at least as large as the number of markets/periods. 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.[9]

Therefore, in most settings, it is impossible to approximate the conditional moments without additional restrictions. This is an important problem for empirical work, since there is a very large number of valid potential instrument functions satisfying Assumption 2, most of which weakly identify model.

2   Exchangeability and the choice of instruments

The previous discussion highlights the role of the instruments in approximating the conditional moment restrictions. In order to derive a class of relevant instrument functions, it is useful to decompose the expectation of the residual function in two components:

$$r_{jt}(\boldsymbol{\theta}) \;\;=\;\; \pi_j(\boldsymbol{x}_t, \boldsymbol{w}_t) - \boldsymbol{x}_{jt}\boldsymbol{\beta}$$

We refer to $\pi_j(\boldsymbol{x}_t, \boldsymbol{w}_t) = E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}\right) | \boldsymbol{x}_t, \boldsymbol{w}_t\right]$ as the reduced-form of the model. Let $\sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t; \boldsymbol{\lambda}) = \boldsymbol{x}_{jt}\boldsymbol{\beta}^0 + \Delta_j\left(\boldsymbol{\lambda}\right) + \xi_{jt}$ denotes the inverse demand function evaluated at parameter λ, expressed as function of the quality gap relative to δ_{jt}:

$$\Delta_{jt}\left(\boldsymbol{\lambda}\right) = \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}) - \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}^0).$$

Using this notation, the reduced form is approximated by the instruments via a linear projection:

$$\begin{aligned}
\pi_{jt}(\boldsymbol{x}_t, \boldsymbol{w}_t; \boldsymbol{\lambda}) &= E\left[\boldsymbol{x}_{jt}\boldsymbol{\beta}^0 + \Delta_j\left(\boldsymbol{\lambda}\right) + \xi_{jt} \big| \boldsymbol{x}_t, \boldsymbol{w}_t\right] \\
&\approx \boldsymbol{x}_{jt}\boldsymbol{\gamma}_1 + A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)\boldsymbol{\gamma}_2 + 0,
\end{aligned} \qquad (7)$$

and $l(\boldsymbol{z}_{jt}; \boldsymbol{\theta}) = \boldsymbol{z}_{jt}\boldsymbol{\gamma} - \boldsymbol{x}_{jt}\boldsymbol{\beta}$ is the linear projection approximating the conditional moments.

The reduced-form parameters $\boldsymbol{\gamma}_2$ determines the strength of the correlation between the excluded instruments, and the quality gap evaluated at parameter $\boldsymbol{\lambda}$. An instrument function is *relevant* if $\boldsymbol{\gamma}_2$ is jointly different from zero in the population. For instance, at $\boldsymbol{\lambda} \to 0 \neq \boldsymbol{\lambda}^0$, the model exhibits IIA preferences, and $\Delta_{jt}\left(\boldsymbol{\lambda}\right)$ is positive

---

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

for products with few substitutes. In contrast, products that are undifferentiated along the price or $\boldsymbol{x}_{jt}^{(2)}$ dimensions are assigned a negative quality gap. When $\boldsymbol{\gamma}_2$ is (jointly) close to zero in the population, the instruments are unable to explain this type of deviations in the assignment of product quality. This leads to a weak identification problem since the linear projection is not statistically different from zero away from the true parameters.

Our objective in this section is to develop a theory-driven approach to constructing instruments that can approximate the reduced-form function $\pi_j(\boldsymbol{x}_t, \boldsymbol{w}_t)$, while circumventing the curse of dimensionality problem. 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 first illustrate this result in the context of a model without endogenous prices. Section 2.2 relaxes this assumption. Section 2.3 introduces two classes of instrument functions that satisfy the symmetry property.

2.1 Exogenous characteristics

With exogenous characteristics, we can write the conditional-moment restrictions as a function of the menu of characteristics $\boldsymbol{x}_t$ (dropping $\boldsymbol{w}_t$):

$$E\left[\rho_j\left(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta}\right) | \boldsymbol{x}_t\right] = \pi_j(\boldsymbol{x}_t; \boldsymbol{\lambda}) - \boldsymbol{x}_{jt}\boldsymbol{\beta} = 0.$$

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 $\boldsymbol{d}_{jt} = (\boldsymbol{d}_{jt,0}, \ldots, \boldsymbol{d}_{jt,j-1}, \boldsymbol{d}_{jt,j+1}, \ldots, \boldsymbol{d}_{jt,J_t})$ 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 $k = 0, \ldots, J_t$ in the market (including the outside good) for a given inside product $j > 0$, and let $\boldsymbol{\omega}_{jt} = \{\boldsymbol{\omega}_{jt,k}\}_{k=0,\ldots,J_t}$. We now have the following result which is proven in Appendix A.

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

$$\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}\right) = f\left(\boldsymbol{\omega}_{jt}; \boldsymbol{\lambda}\right) + C_t(\boldsymbol{\lambda}), \quad j = 1, \ldots, J_t \tag{8}$$

*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 (8).[10] These two normalizations imply that the demand function for product j is a fully exchangeable function of the structure of the market *relative* to product j: $\boldsymbol{m}_{jt} = \left\{ (\boldsymbol{d}_{jt,k}^{(2)}, \tau_{0t}) \right\}_{k=0,\ldots,J_t}$. 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.[11] The second implication is that $f(\cdot)$ is a *symmetric* function of the states of the competing products. Proposition 1 can be viewed as an extentension of the the partial-exchangeability result obtained in Pakes (1994) to reduce the dimensionality of equilibrium strategies in differentiated product markets (e.g. investment and pricing).

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 $\{\xi_{1t}, \ldots, \xi_{J_t,t}\}$ is exchangeable, the conditional expectation of the inverse-demand is a symmetric function of the matrix of characteristic differences:*

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

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 cor-

---

[10]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.

[11]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.

responds 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 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.

Relevant instrument function: Differentiation IVs

Proposition 2 is important because it allows us to allows us to use low order approximations of the reduced form without the problems of curse of dimensionality discussed above. A relevant instrument is a function that satisfies the symmetry property, while summarizing the distribution of characteristics differences relative to product j. Since these functions measure the degree of differentiation, we label them *Differentiation IVs*.

To understand why the symmetry of the reduced-form solves the curse of dimensionality problem, 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: $d_{jt,k} = x_{kt} - x_{jt}$. If we use a monomial basis to approximate π, the first order polynomial 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, $g(\boldsymbol{d}_{jt,-j}; \boldsymbol{\lambda}) = g(\boldsymbol{d}_{jt,\iota(-j)}; \boldsymbol{\lambda})$, 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 little variation across products within a market, especially when the number of products per market is large. We therefore focus on the sum of square of characteristic difference as the relevant term to construct the instrument function (Donald et al. 2008). Importantly, this instrument has an economic interpretation since it measures the (square of) Euclidian distance of product j along dimension x.

Another common class of basis functions that exploit the symmetry of the function is the spline. Consider for instance a linear B-Spline. The basis function over sequence of knots $c_1, \ldots, c_M$ is:

$$B_{jt,j'}^m \begin{cases} 1 & \text{If } c_m < d_{jt,j'} < c_{m+1} \\ 0 & \text{Else.} \end{cases} \tag{9}$$

This would lead to the following approximation function:

$$g(\boldsymbol{d}_{jt}; \boldsymbol{\lambda}) \approx \sum_{j' \neq j} \sum_{m=1}^{M} B_{jt,j'}^m \gamma_{j',m} = \sum_{m=1}^{M} N_{jt,j'}^m \gamma_m \tag{10}$$

Where the last equality imposed the symmetry of the function, and $N_{jt,j'}^m = \sum_{j' \neq j} B_{jt,j'}^m$ is the number of rival products located with the neighborhood (c_m, c_{m+1}) of product j. Again this instrument function correspond to a commonly used measure of product differentiation in the literature.

In summary, Proposition 1 and 2 solve 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.[12]

---

[12]See Altonji and Matzkin (2005) and Farias, Saure, and Weintraub (2012) for a related uses of symmetric functions.

14

2.2 Endogenous attributes

Incorporating endogenous characteristics, such as prices or advertising, 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 in equation (5) to identify the model is no longer satisfied at $\boldsymbol{\theta}^0$:

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

The challenge 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 monopoly 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.

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. We proceed in two steps.

First, let $\hat{p}_{jt} \approx E(p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t)$ denotes an estimate of the reduced-form 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, $\hat{p}$ can be estimated using 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:

$$
\begin{aligned}
E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda}\right)\big|\boldsymbol{x}_t, \boldsymbol{w}_t\right] &\approx E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \hat{\boldsymbol{p}}_t; \boldsymbol{\lambda}\right)\big|\boldsymbol{x}_t, \hat{\boldsymbol{p}}_t\right] \\
&= g\left(\boldsymbol{d}_{jt}, \boldsymbol{d}_{jt}^{\hat{p}}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda})
\end{aligned}
\tag{11}
$$

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 $\{\boldsymbol{d}, \boldsymbol{d}_{jt}^{\hat{p}}\}$. In what follows, we will construct instrument functions that exploit the symmetry of this function.

2.3 Practical implications

In this section, we propose two classes of low-order instrument functions that satisfy the symmetry property, and exhibit variation across products. Depending on the data and richness of the model, more or less complicated functions can be used. As we alluded to earlier, the key is to select instrument functions that predict differences in the inverse-demand of products (within a market) as we vary the main parameters of the model.

Our first example uses the leading terms of a second-order symmetric polynomial

basis function (focussing only on the binary interaction terms):

$$
A_j(\boldsymbol{x}_t, \boldsymbol{w}_t) = \begin{cases} \boldsymbol{w}_{jt} & \text{Price IVs} \\ \sum_{j' \neq j} \left(d^{\hat{p}}_{jt,j'} \right)^2, & \forall k & \text{Isolation } j \text{ in price} \\ \sum_{j' \neq j} \left(d^{k}_{jt,j'} \right)^2, & \forall k & \text{Isolation } j \text{ in } x_k \\ \sum_{j' \neq j} d^{k}_{jt,j'} \times d^{l}_{jt,'}, & \forall k \neq l & \text{Interaction: } k \text{ and } l \end{cases} \tag{12}
$$

where $d^{k}_{jt,j'} = x_{j't,k} - x_{jt,k}$ measures the difference between product j and j' along dimension k. Note that one dimension refers to the price. The interaction terms for which $l \neq k$, capture the covariance between two dimensions of differentiation, while the sum of square measure the isolation of products in the characteristic space.

Our second instrument function is based on the linear B-spline example discussed above; 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, which is why to focus only on the fist Spline segment. 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:

$$
A_j(\boldsymbol{x}_t, \boldsymbol{w}_t) = \begin{cases} \boldsymbol{w}_{jt} & \text{Price IVs} \\ \sum_{j' \neq j} 1 \left(|d^{\hat{p}}_{jt,j'}| < \kappa^{\hat{p}} \right), & \forall k & \text{Isolation of } j \text{ in price} \\ \sum_{j' \neq j} 1 \left(|d^{l}_{jt,j'}| < \kappa^{k} \right), & \forall k & \text{Isolation } j \text{ in } x_k \\ \sum_{j' \neq j} 1 \left(|d^{l}_{jt,j'}| < \kappa^{k} \right) \times d^{l}_{jt,'}, & \forall k \neq l & \text{Interaction: } k \text{ and } l \end{cases} \tag{13}
$$

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. Moreover, additional neighborhoods can

be constructed to impose additional restrictions on the model (e.g. $0 < |d_{jt,j'}| \leq \kappa_1$, $\kappa_1 < |d_{jt,j'}| \leq \kappa_2$, etc.)

The two formulations of the Differentiation IVs in equations (33) and (13) 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. $\boldsymbol{x}_{jt}^{(2)} = \boldsymbol{x}_j^{(2)}$ $j = 1, \ldots, 25$), while prices vary both at the product and the market level. Using exogenous variation in prices across markets, we can construct instruments as follows:

$$A_j(\boldsymbol{x}_t, \boldsymbol{w}_t) = \left\{ \boldsymbol{w}_{jt}, \sum_{j' \neq j} 1\left(|d_{j,j'}^1| < \kappa^1\right) d_{jt,j'}^{\hat{p}}, \ldots, \sum_{j' \neq j} 1\left(|d_{j,j'}^K| < \kappa^K\right) d_{jt,j'}^{\hat{p}} \right\}. \quad (14)$$

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.

Incorporating demographic characteristics in the instrument function is straight-forward. Appendix A.3 provides an example in which a random coefficient is an additive function of income (y_{it}). If the distribution of income can be standardized (i.e. $y_{it} = m_t + \text{sd}_t e_{it}$, where $e_{it} \sim \psi(e_{it})$), the reduced-form of the model can be written as a symmetric function of characteristics differences and moments of the demographic distribution:

$$\pi_{jt}(\boldsymbol{x}_t; \boldsymbol{\lambda}) = g\left(\boldsymbol{d}_{jt}, \text{sd}_t \cdot \boldsymbol{d}_{jt}^{(2)}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda}) \quad (15)$$

In addition, the average utility index (δ_{jt}) includes an interaction variable: $m_t \cdot x_{jt}^{(2)}$. This variable can be used as an excluded instrument (as in Miller and Weinberg (2017)). Therefore, in this example the instrument vector includes the full vector of products' own characteristics (including $m_t \cdot x_{jt}^{(2)}$), as well as moments of the distribution of characteristics differences interacted with the standard-deviation of y_{it} in market t.

The takeaway is that market-specific moments of the distribution of demographics should enter the instrument function as interaction terms with product characteristics and differentiation measures, rather than as stand-alone variables.[13]

---

[13]See Romeo (2010) for a similar argument and simulation results showing the importance of ac-

How does this formulation 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 (or markups), 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 (13). 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 Additional implications for estimation and testing

In this section we discuss three additional implications of the exchangeability results from Section 2 for the identification and estimation of the model. We start by discussing approximation to the optimal instruments, and the non-parametric estimation of the model. We then propose a test of instrument relevance that can be used in the case of the multinomial Logit model with random coefficients.

3.1 Optimal IV approximation

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. The following corollary establishes a direct connection between product differentiation and optimal instruments.

---

counting for interactions between product characteristics and the mean of demographic attributes in the instrument vector.

Corollary 1. *If the distribution of $\{\xi_{1t}, \ldots, \xi_{J_t,t}\}$ 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. Therefore, 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 non-parametric regressions as discussed in Newey (1993), or the heuristic approximation discussed in Berry, Levinsohn, and Pakes (1999) and Reynaert and Verboven (2013).

This later approach relies on the following instrument function:

$$E\left[\left.\frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right| \boldsymbol{x}_t, \boldsymbol{w}_t\right] \approx \left.\frac{\partial \rho_j(\boldsymbol{s}_t, \hat{\boldsymbol{p}}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\xi_{jt}=0, \forall j, t} = A_j(\boldsymbol{x}_t, \boldsymbol{w}_t | \boldsymbol{\theta}). \quad (16)$$

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 $A_j(\boldsymbol{x}_t, \boldsymbol{w}_t | \boldsymbol{\theta}^1)$ and estimate $\hat{\boldsymbol{\theta}}$ by GMM.

Although this approach works well in practice (Reynaert and Verboven 2013), the performance of the estimator depends on using strong IVs in the first-stage estimates. This suggests a strong complementarity between the two approaches as discussed in Conlon and Gortmaker (2019). 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.

3.2 Non-parametric estimation

Another implication of Propositions 1 and 2 is that it is feasible to construct a consistent semi-parametric estimator of the linear-in-characteristics random-coefficient model. To

see this, consider the following quasi-linear indirect utility:

$$u_{ijt} = x_{jt}\beta - p_{jt} + \xi_{jt} + x_{jt}^{(2)}\lambda_i + \epsilon_{ijt} \tag{17}$$

where $(\epsilon_i, \lambda_i) \sim F(\epsilon_i, \lambda_i)$, and $\delta_{jt} = x_{jt}\beta - p_{jt} + \xi_{jt}$. We treat the distribution of heterogeneity as non-parametric, and focus on the semi-parametric estimation the demand model.

The inverse demand function of this model can be written as:

$$p_{jt} = x_{jt}\beta - \sigma_j^{-1}\left(s_t, x^{(2)}\right) + \xi_{jt}.$$

This corresponds to a partial linear regression model with an endogenous non-parametric function. This is one of the two examples studied in Ai and Chen (2003), and the conditional moment restriction can be used to construct a Sieve minimum distance (SMD) estimate for β and b. This requires approximating the structural (inverse-demand) and reduced-form equations using two linear Sieves.[14] This estimator breaks the curse of dimensionality in the number of products by using basis functions that satisfy the exchangeability conditions implied by Proposition 1 and 2.

Compiani (2019) demonstrates that Bernstein polynomial can be used in this context. In small sample, the inverse demand can also be approximated using the "distance-metric" approach proposed by Pinkse, Slade, and Brett (2002).

3.3 Testing for instrument relevance

One important challenge when evaluating the strength of the excluded instruments is that the reduced-form function depends on an unknown parameter vector $\boldsymbol{\lambda}$. 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. This test relies on the assumption that ϵ_{ijt} is distributed according to a T1EV distribution.

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 multi-

---

[14]When the price enters the inverse-demand non-linearly, the same reduced-form representation can be used, but it would rely on the heuristic discussed in Section 2.2.

nomial logit model. A similar exclusion restriction can be tested using the aggregate inverse-demand function at $\boldsymbol{\lambda} = 0$ (Berry 1994):

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

Where, as in Section 1.2, $\Delta_j(\cdot)$ measures deviations from IIA in the true model (i.e. quality-gap). We construct an "IIA-regression" by taking expectation of shares and prices, 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 1 (IIA-test). *If Assumption 2 is valid, the IIA hypothesis can be tested by estimating the following regression:*

$$
\begin{aligned}
E[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t, \boldsymbol{w}_t] &= \boldsymbol{x}_{jt}\boldsymbol{\beta} + \alpha E\left[p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t\right] + E\left[\Delta_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}, \boldsymbol{p}_t; \boldsymbol{\lambda} = 0)|\boldsymbol{x}_t\right] + 0 \\
&\approx A_j(\boldsymbol{x}_t, \boldsymbol{w}_t)\boldsymbol{\gamma} = \boldsymbol{x}_{jt}\boldsymbol{\gamma}_0 + \gamma_p \hat{p}_{jt} + A_j^{-w}(\boldsymbol{x}_t, \boldsymbol{w}_t)\boldsymbol{\gamma}_2, \quad (18)
\end{aligned}
$$

where $A_j^{-w}(\boldsymbol{x}_t, \boldsymbol{w}_t)$ is a partition of the instrument vector that excludes product j's price IV ($\boldsymbol{w}_{jt}$). The null hypothesis of IIA preferences correspond to: $H_0 : \hat{\boldsymbol{\gamma}}_2 = 0$.[15]

Note that the IIA regression corresponds to the reduced-form of the model evaluated at $\boldsymbol{\lambda} = 0$, and therefore suffers from a curse of dimensionality problem. Using Proposition 2 we can express this conditional expectation as a symmetric function of the distribution of characteristic differences. Corollary 2 formalizes this result.

Corollary 2. *If the distribution of $\{\xi_{1t}, \ldots, \xi_{J_t,t}\}$ is exchangeable and the price heuristic defined in equation (11) provides a good approximation to the reduced-form, the IIA regression can be written as a symmetric function of the matrix of characteristic differences:*

$$
\begin{aligned}
E\left[\ln s_{jt}/s_{0t}|\boldsymbol{x}_t, \boldsymbol{w}_t\right] &= \boldsymbol{x}_{jt}\boldsymbol{\beta} + \alpha\hat{p}_{jt} + E\left[\Delta_j(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda} = 0)|\boldsymbol{x}_t, \boldsymbol{w}_t\right] \\
&\approx \boldsymbol{x}_{jt}\boldsymbol{\gamma}_1 + \gamma_p \hat{p}_{jt} + h(\boldsymbol{d}_{jt}^x, \boldsymbol{d}_{jt}^{\hat{p}}) + h_t^0 \quad (19)
\end{aligned}
$$

---

[15]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 $E\left[\Delta_j(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^0)|\boldsymbol{x}_t, \boldsymbol{w}_t\right] \approx A_j^{-x}(\boldsymbol{x}_t, \boldsymbol{w}_t)\boldsymbol{\gamma}_1$, 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 ($\boldsymbol{\beta}^0$).

where h_t^0 is a market-specific intercept.

The key implication of this corollary is that the IIA hypothesis can be tested by measuring the strength of the correlation between market shares and measures of product differentiation. There exists several ways of testing the exclusion restriction that product differentiation is independent of demand.

Perhaps the most intuitive approach is to estimate the multinomial Logit model by 2SLS while controlling for the Differentiation IVs:

$$\ln s_{jt}/s_{0t} \quad = \quad \boldsymbol{x}_{jt}\hat{\boldsymbol{\gamma}}_1 + \hat{\gamma}_p p_{jt} + A_j^{-w}(\boldsymbol{x}_t, \boldsymbol{w}_t)\boldsymbol{\gamma}_2 + \text{error},$$

where the cost shifters $\boldsymbol{w}_{jt}$ are used as an excluded instrument for price. In this case, the IIA-test corresponds to the null hypothesis $H_0 : \hat{\boldsymbol{\gamma}}_2 = 0$. This test can be implemented without solving for the non-linear demand function using standard Wald or F tests, and is therefore robust to mis-specification of the demand model. As we show below, the sign of the coefficient $\boldsymbol{\gamma}_2$ is informative about the nature substitution patterns under the true underlying model. This regression can therefore be used to guide the specification of the random-coefficient model (e.g. which characteristic should be interacted with a random-coefficient).

An equivalent approach to test the IIA hypothesis is to estimate a mis-specified model using both $\boldsymbol{w}_{jt}$ and A_j^{-w} as instruments. In this case, the IIA-test corresponds to an over-identification test measuring the validity of the exclusion restrictions formed by the price instruments and the Differentiation IVs. This can be done using the Sargan-Hansen J-test. We discuss both approaches in the application below.

In both cases a rejection of the null hypothesis implies that the excluded instruments detect statistically significant differences between the true and the multinomial Logit inverse-demand functions. We interpret this as a measure of the relevance of the instrument function. The predicted value $A_j^{-w}(\boldsymbol{x}_t, \boldsymbol{w}_t)\hat{\gamma}_2$ quantifies the expected deviations from IIA under the true model by estimating a (potentially) mis-specified model. As we discussed in Section 1.2, failure to reject the IIA hypothesis when $\boldsymbol{\lambda}^0 \neq 0$ is consistent with weak-identification.

Of course, the instruments can be "strong" 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 ($\boldsymbol{\lambda}^0 = 0$). Alternatively, if the the data does not exhibit enough variation in product characteristics, either within or

across markets, it might not be feasible to find statistically significant deviations from IIA. In this case, the previous discussion suggests that the non-linear model will be weakly identified, and inference methods robust to weak identification should be used.

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.

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. We perform all numerical integrations by discretizing the distribution of the random coefficients, and use the same grid and weights in the monte-carlo and estimation steps; therefore avoiding any mis-specification due to simulation errors. We provide a pseudo-code description of our approach in Appendix B.2. We also provide sample Phython and Ox codes on our website.[16]

4.1 Independent random-coefficients

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 then illustrate how the Differentiation IVs can alleviate the problem.

Consider the following IID random-coefficient model with **exogenous** characteris-

---

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

Figure 1: IIA test with weak and strong instruments

(a) Weak IV: Sum

(b) Strong IV: Euclidian distance
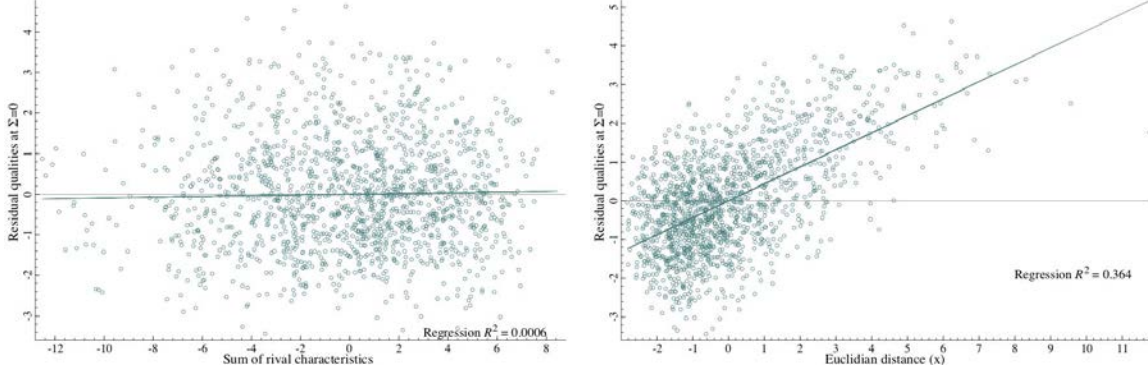
tics:

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + \sum_{k=1}^{K_2} (\beta_{2,k} + \lambda_k \eta_{ik}) \cdot x_{jt,k}^{(2)} + \xi_{jt} + \epsilon_{ijt} \tag{20}$$

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 $T = 100$.

We compare the performance of three instrument functions:

Sum of characteristics IV: $\qquad A_j(\boldsymbol{x}_t) = \left\{ \boldsymbol{x}_{jt}, \sum_{j' \neq j}^{J} \boldsymbol{x}_{j',t} \right\}$

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

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

where K is the number of characteristics (excluding the intercept), and sd_k is the standard-deviation of $x_{jt,k}$.

Figure 1a illustrates the IIA-test graphically in the single-dimensional model with weak instruments. Each dot represents a product/market combination, and the line

corresponds to a linear regression of $\hat{r}_{jt}$ on the instrument.[17] 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. In contrast, the performances of the Differentiation IVs is nearly identical when using a large number of products, since the IVs exhibit substantial variation across products within the same market even when J_t grows large.

Figure 1b illustrates the correlation between the residual function at $\lambda = 0$ (Logit) and the *Euclidian* distance of product j (i.e. strong instrument) for same model.[18] 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. Importantly, the sign of the correlation is an important indicator of the model specification. A positive coefficient on distance indicates that the availability of close substitutes reduces the probability of buying j, or equivalently that products with similar attributes are close substitutes. In contrast, a negative coefficient would be inconsistent with the random-coefficient model of demand (i.e. presence of similar rivals *increase* demand). If that was the cause, it would likely indicate a violation of the conditional independence assumption.[19] Misspecification of this type would lead to an estimate of the random-coefficient parameter equal to zero (corner), since the model cannot rationalize this reduced-form relationship (see Houde (2012) for a discussion of mis-specification in the context of a model of spatial differentiation). It is therefore important in applied work to measure the strength and the sign of the

---

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

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

[19]For instance, this correlation can be explained by an (unobserved) increase in the willingness-to-pay for a certain attribute that cause entry of products with the same attributes.

reduced-form correlation between differentiation and demand prior to estimating the structural model.

This positive relationship between differentiation (or distance) and the inverse demand at $\lambda = 0$ is captured by the differentiation instrument used in Figure 1b. In other words, 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 to products that are relatively isolated, and low quality to products with many substitutes. A clear violation of the moment conditions.

The results of 1,000 Monte-Carlo replications with weak instruments are summarized in Table 1a. We estimate the log of λ_k in equation (20), instead of λ_k directly, to account for the strictly positive support of the parameter space. The table reports the mean bias and RMSE for the transformed parameters, averaged across parameters ($\lambda_k^0 = 4$ for all λ_k). Table 2 summarizes the small-sample performance for the two IVs across all specifications, and calculates the average asymptotic standard-errors. Tables 3a and 3b summarize the full set of simulation results, including the weak identification and IIA tests, and the local minimum statistics.

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. The bottom panel of Table 1a reports the results of the two tests.

We test the null-hypothesis of IIA preferences by testing the joint null hypothesis that $\hat{\boldsymbol{\gamma}}_1 = 0$. As the figures suggest, 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.[20]

Next, we look at the finite-sample performance of the GMM estimator under weak identification. Note that in 8.4% of the samples, 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)). Weak

---

[20]We use a 10% confidence level calculated using Stock-Yogo critical values to calculate the rejection probabilities (Stock and Yogo (2005)).

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

(a) Weak instruments

	$K_2 = 1$		$K_2 = 2$		$K_2 = 3$		$K_2 = 4$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
λ_k	0.136	2.643	0.06	2.46	0.09	2.25	0.12	2.301
1(Local-min)	0.189		0.514		0.594		0.661	
Range(J-stat p-value)	0.167		0.189		0.212		0.210	
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	

(b) Strong instruments

	bias	rmse	bias	rmse	bias	rmse	bias	rmse
	$K_2 = 1$		$K_2 = 2$		$K_2 = 3$		$K_2 = 4$	
λ_k	0.002	0.122	-0.003	0.1275	0.0015	0.129	-0.003	0.142
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	

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. The bias and RMSE are averaged across parameters: $\lambda_k = \{\lambda_1, \ldots, \lambda_K\}$.

instruments imply that the normal distribution is a poor approximation of the finite-sample distribution of the parameter estimates, and causes the presence of frequent outliers (leading to corner solutions).

Another consequence of weak instruments is the lack of precisions in the 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 is poor across all four specifications, and remains constant as we increase the complexity of the model.

A third 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

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.

1(Local-min) is equal to one if the algorithm converged to more than one solution.

Using this procedure, we find large number of "local minima". The frequency of this 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 and flat GMM objective function. This makes it difficult for Newton and quasi-Newton algorithms to find the global minimum when instruments are weak.

The next row of Table 1a illustrates the magnitude of the differences between these different local solutions. 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).

We now turn to the simulation results obtained with the quadratic Differentiation IVs. We obtain similar results with the Local Differentiation IVs. Table 1b presents the average bias and RMSE across parameter (λ_k). 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 the same solution

Table 3: Monte-Carlo simulation results for exogenous characteristics model with strong instruments

(a) Differentiation IV: Quadratic

	bias	rmse	bias	rmse	bias	rmse	bias	rmse
	$K_2 = 1$		$K_2 = 2$		$K_2 = 3$		$K_2 = 4$	
λ_k	0.002	0.122	-0.003	0.1275	0.0015	0.129	-0.003	0.142
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
λ_k	0.002	0.126	-0.0003	0.13	0.002	0.133	-0.003	0.148
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	

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 1a.

We also find that minimal loss in precision from adding random-coefficients. 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.

Finally, Table 4 compares the bias and precision of the estimates between non-linear least-square and GMM. The first two specifications reproduces the bias and RMSE results from the specification with four random-coefficients. The non-linear least-square (NLS) estimates correspond to a specification where the parameters are obtained by minimizing the sum of square residuals. This specification leads to biased estimates

Table 4: Monte-Carlo simulation results with NLS and GMM

Parameter	Diff. IV (quad.) Bias	SE	Sum Characteristics Bias	SE	NLS Bias	SE
λ_1	-0.003	0.142	0.218	2.348	.227	.0238
λ_2	-0.004	0.141	0.099	2.297	.225	.0239
λ_3	-0.001	0.137	0.113	2.378	.226	.0238
λ_4	-0.005	0.146	-0.075	2.207	.225	.0239

arising because of the simultaneity of market shares. Column (3) shows that this leads to an average downward bias of roughly 10% (i.e. $\lambda_k^0 = 2$). The main advantage of least-square is the precision of the estimates. The asymptotic standard errors are equal to 0.023; significantly smaller than the RMSE with GMM. This highlights the tradeoff between bias and precision induced by the choice of instruments. With weak instruments researchers should put a lot more importance of least-square results, given the large mean-square error associated with GMM.

4.2 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 13 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

Table 5: Simulation results for the correlated random-coefficient model

		$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$	$\Sigma_{\cdot,4}$
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 13 in the Appendix presents the variance-covariance matrix of ν_{ik}. Number of simulations: 1,000.

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 5. The top panel reports the average estimated parameters (transformed) of the variance-covariance matrix, the middle panel reports the RMSE associated with each parameter, and the bottom panel reports the averages of the IIA-test and the Cragg-Donald rank test statistics. Both tests confirm 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.

---

[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.

Figure 2: Distribution of estimated price random-coefficient parameter

Dash vertical line = True parameter value
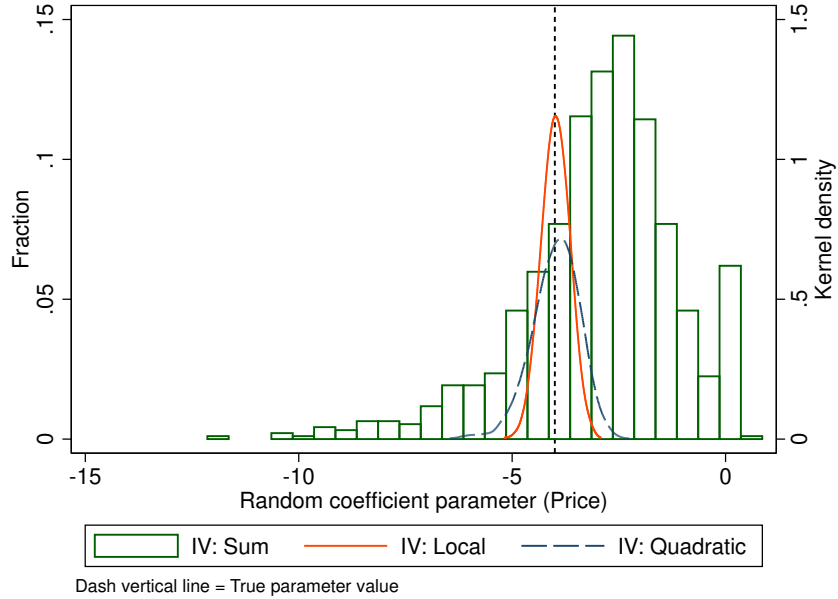
It is worth noting that this specification is substantially richer than any random-coefficient model that has previously been studied with aggregate data by researchers, both in empirical applications and Monte-Carlo simulations. Although we obtain these results in a "controlled" environment, this result 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.

4.3 Endogenous prices

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

$$u_{ijt} = \beta_0 + \beta_1 x_{jt}^{(1)} + (\beta_p + \lambda_p \nu_i) \cdot p_{jt} + \xi_{jt} + \epsilon_{jt}, \quad j, = 1, \dots, 15 \text{ and } t = 1, \dots, 100. \quad (21)$$

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

---

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

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

(a) Distribution of parameter estimates

	(1) True	(2) Diff. IV = Local			(3) Diff. IV = Quadratic			(4) 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) IV = Local	(2) IV=Quadratic	(3) 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.

following vector of first-order conditions:

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

The marginal cost, $c_{jt} = \gamma_0 + x_{jt}^{(1)}\gamma_x + w_{jt}$, is constant, and the cost-shock w_{jt} is observed by the econometrician. We use this variable below to construct a price instrument. The data is generated by finding a solution to equation (22) 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

---

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

regression of p_{jt} on the exogenous characteristic and the cost shifter w_{jt}:

$$\hat{p}_{jt} = \hat{\pi}_0 + \hat{\pi}_1 x_{jt}^{(1)} + \hat{\pi}_2 w_{jt}. \tag{22}$$

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 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:

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

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

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 2 and Table 6. In addition to the two sets of instruments defined above, we also report the results using the "sum of rival characteristics" in order to illustrate effect weak instruments. The bottom panel reports the results of the weak identification and IIA tests.

Table 6b confirms the weak identification results obtained in the models with exogenous characteristics. 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).

To illustrate the dual role of the instruments in this context, Table 6b 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 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 6a 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 2 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.

4.4 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 2 either because of the endogenous selection of products, and/or because of a contemporaneous correlation between ξ_{jt} and the attributes of own

---

[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.

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.

Assumption 4. *The change in the unobserved quality of products is mean zero con-*

---

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

ditional 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:

$$m_n(\lambda) = \frac{1}{n}\sum_m\sum_j \left[\rho_j(\boldsymbol{s}_{m1}, \boldsymbol{x}_{m1}; \lambda) - \rho_j(\boldsymbol{s}_{m0}, \boldsymbol{x}_{m0}; \lambda)\right] \cdot \boldsymbol{z}_{jm} = \Delta\boldsymbol{\rho}(\lambda)^T\boldsymbol{z}/n$$
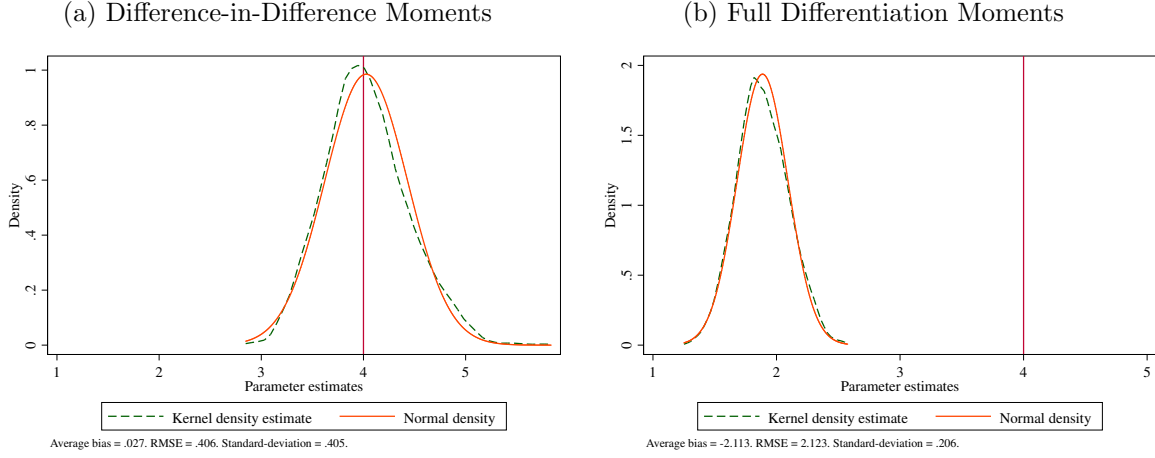
where n is the number of unique market/product observations. Using this specification, the structural parameters of the model are identified solely from the quasi-experimental variation. In particular, the reduced-form is approximated by a difference-in-difference regression, in which the "control" group is defined as the set of products located relatively far from the exogenous new entrant.

Figure 3 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 (3a), or the full "Differentiation IVs" moments (3b).[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 3b 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

---

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

(a) Difference-in-Difference Moments

(b) Full Differentiation Moments

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

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 $\bar{\xi}_{jm} = -0.25 \left(\mathrm{ED}_{jm} - \overline{\mathrm{ED}}_m \right) + \zeta_{jm}, \zeta_{jm} \sim N(0, 0.5)$ 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, $T = 2$.

a small disutility from distance. Figure 3b illustrates that the difference-in-difference moment conditions eliminate this bias. The distribution is centered around $\lambda^0 = 4$, and the average bias is less than 1% of the parameter value.

Comparing the two distributions, it is important to note that by exploiting solely the variation created by the entry of a new product, the difference-in-difference GMM estimator is less precise, and the distribution of $\hat{\lambda}$ is less well approximated by the normal density than the specification that uses the larger set of instruments. In Figure 3b the p-value associated with Shapiro-Wilk normal test is 11%, compared to less than 1% in Figure 3a. 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)).

4.5 Comparison with other approaches

Finally, we conclude this section 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)):

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

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:

$$
E\left[\frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|\boldsymbol{x}_t\right] \approx \frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\xi_{jt}=0, \forall j, t} = \tilde{A}_j(\boldsymbol{x}_{jt}|\boldsymbol{\theta}). \tag{23}
$$

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:

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

$$\tag{24}$$

---

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

Table 7: Optimal IV approximation with alternative initial parameter values

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

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.

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:

$$\text{Normal RC:} \qquad u_{ijt} = \delta_{jt} + \lambda \eta_i x_{jt}^{(2)} + \epsilon_{ijt}$$

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

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

Table 7 summarizes the results of 1,000 Monte-Carlo replication simulations. The first five rows correspond to different values of the initial parameter used to evaluate the Jacobian. In both 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 7 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 4.2. To illustrate the importance of using consistent estimates in the first-stage, we implement the optimal IV approximation using pseudo-random

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

Choleski matrix	True	Opt. IV: $\boldsymbol{\theta}^1 \sim N(0,1)$			Opt. IV: $\boldsymbol{\theta}^1 \sim N(0,4)$			Diff. IV: Quad.		
		bias	rmse	se	bias	rmse	se	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

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 8. 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 4.3. Table 9 summarizes the results. The top-panel

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

		True	Diff. IV = Local			Diff. IV = Quadratic			Diff. IV = Sum		
			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

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 (23). 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 7) 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.

5   Application: Demand for new cars

In this section, we apply our testing and identification strategy to study demand for new cars between 1971 and 1990. Our objective is twofold. First, we use this application to illustrate our suggested approach to select instruments, and conduct inference. Second, we use the car application to demonstrate that it is feasible to identify the model solely using demand restrictions, and without relying on external moment conditions coming from a pricing model (as in Berry et al. (1995)) or survey data (Petrin (2002), and Berry et al. (2004)).

We use the panel data-set analyzed first by Berry et al. (1995). We enrich the model used in Berry et al. (1995) by controlling for market and brand fixed-effects (as in Nevo (2001)). In particular, we consider the following payoff function:

$$u_{ijt} = x_{jt}\beta + \beta_p p_{jt} + \text{Make FE} + \text{Year FE} + \xi_{jt} + \lambda_p p_{jt}/y_i + \sum_{k=1}^{K} \lambda_k x_{jt,k} \eta_{ik} + \epsilon_{ijt} \quad (25)$$

where $y_i \sim LN(\mu_t^y, \sigma_t^y)$, and $\eta_{ik} \sim N(0,1)$. We use 25 brands to construct our make variable; grouping smaller European and Asian brands together. The distribution of household income is approximated by a log-normal distribution with time-varying location and spread parameters. The other random coefficients assumed to be IID normally distributed. We use 500 Halton draws to integrate the random coefficients (see Train 2009).

We use a sequential approach to determine the set of instruments and the random coefficient specification. We first conduct a series of reduced-form specification tests to evaluate the the strength of price and differentiation instruments, and then estimate the model by non-linear GMM. Note that we use the reduced-form specification tests to guide the specification of the random-coefficient model estimated in the second step.

Table 10: Multinomial Logit demand for cars and IIA tests

VARIABLES	(1)	(2)	(3)	(4)	(5)
Price (x1000)	-0.153***	-0.152***	-0.127***	-0.009	0.000
	(0.040)	(0.039)	(0.034)	(0.012)	(0.011)
Distance: Pred. price	0.016***	0.016***	0.014***		
	(0.004)	(0.004)	(0.004)		
Distance: Nb. doors		0.027**	0.022		
		(0.014)	(0.013)		
# same attribute: Air			0.009**		
			(0.004)		
Observations	2,217	2,217	2,217	2,217	2,217
Weak IV (K-P)	34.82	35.88	57.79	59.49	37.76
J-test	0.891	0.894	3.202	19.23	29.96
J-test (p-value)	0.345	0.344	0.0736	0.000709	0.000870
Degree of over-id.	1	1	1	4	10
Joint test: $\gamma_k = 0$ (chi2)	16.19	17.41	17.56		
Joint test (p-value)	5.74e-05	0.000166	0.000542		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We approximate the reduced-form model of price using cost shifters w_{jt}:

$$p_{jt} = x_{jt}\pi_1 + w_{jt}\pi_2 + \text{Make FE} + \text{Year FE} + u_{jt}. \qquad (26)$$

Let $\hat{p}_{jt}$ denotes the OLS predicted price. We use this variable to measure the quality differentiation of cars in the market.

We use two cost-shifters as instrument for prices. We use the weight of the car interacted with ratio of steel and bauxite prices to approximate the cost differential associated with this strategy. Our second price instrument is an indicator variable equal to one after a model started to be produced in the U.S.. By 1990, 10% of cars sales were produced by foreign brands established in the US.[28] The first-stage F-test associated with these two instruments is reported in the first three columns of Table 10. Depending on the controls, the Kleibergen-Paap statistics ranges from 34 to 57, confirming that the two variables are strong predictors of prices.

Next we analyze the reduced-form model by estimating the multinomial Logit model (i.e. $\mu_{ij} = 0$). Recall that we can test the null hypothesis of IIA preferences by

---

[28]We use the following list of off-shored models and dates: VW Rabbit (1979), Honda Accord (1981), VW Golf (1985), Honda Civic (1986), VW Jetta (1986), and Toyota Camry (1989).

evaluating the validity of the exclusion restrictions implied by the assumption that the characteristics of rival products are independent of the unobserved product attributes ξ_{jt} (conditional on own characteristics fixed-effects). We implement two versions of this test.

First, we estimate the model under the null while controlling for differentiation variables (denoted by z_{jt}^{dist}):

$$\ln s_{jt}/s_{0t} = \beta_p p_{jt} + x_{jt}\boldsymbol{\beta}_x + z_{jt}^{\text{dist}}\boldsymbol{\gamma} + \text{Make FE} + \text{Year FE} + e_{jt}. \qquad (27)$$

This equation is estimated by (linear) GMM using w_{jt} as instruments for price. This approach is feasible only because the two IVs (material price and offshoring) are constructed independently of our measures of differentiation. Price instruments that are constructed from the characteristics of rival firms would be invalid if the true model is not Logit, leading to a biased estimate of β_p.

We use measures of product isolation among selected car attributes to construct z_{jt}^{dist}. For continuous variables, we measure differentiation along dimension k using the Euclidian distance.[29] For price, distance is calculated using $\hat{p}_{jt}$. For discrete attributes (e.g. domestic or air) we use the number of products with the same attribute as a measure of differentiation.

In this application we focus on three attributes: price, number of doors, and air conditioning. We use these three variables as Differentiation IVs, and we assign random-coefficients to each characteristic. Other attributes were analyzed as well (e.g. mileage, HP, car size, etc), but we find that differentiation along other dimensions are not correlated with the inverse-demand under Logit.

Columns (1)-(3) in Table 10 test the hypothesis that $\boldsymbol{\gamma} = 0$ using three sets of the differentiation variables. The IIA-tests are reported at the the bottom of the table. The test evaluates the joint hypothesis that $\hat{\boldsymbol{\gamma}}^{gmm} = 0$. In this case the p-value measures the ability of the instruments to identify deviations from IIA (i.e. lower p-value = stronger instruments).

We find strong evidence that product isolation along the price dimension leads to higher market shares; a clear violation of the IIA hypothesis. The positive reduced-form coefficient λ_p is consistent with a quality-ladder model of demand, in which the price is inversely proportional to quality (as in Bresnahan (1987)). Differentiation in

---

[29]We obtain similar results using the number of nearby products. However the Euclidian distance is more strongly correlated with demand in this application.

Table 11: Mixed-logit car demand results

VARIABLES	NLS		GMM		
	(1)	(2)	(3)	(4)	(5)
λ_p	−0.722*	−2.267**	−2.525**	−0.768**	−0.854**
	(0.38)	(1.00)	(1.17)	(0.31)	(0.34)
λ_{door}	0.318*	0.594***	0.631***	0.458***	0.456***
	(0.19)	(0.19)	(0.18)	(0.17)	(0.15)
λ_{air}	0.733		1.833***		0.977
	(1.14)		(0.66)		(0.72)
$\lambda_{\text{Intercept}}$	1.817			7.167**	7.563
	(2.63)			(3.36)	(6.20)
β_{price}	0.052**	0.109***	0.143***	0.064***	0.059***
	(0.02)	(0.03)	(0.05)	(0.01)	(0.01)
Overid. test (J)		0.221	0.119	12.599	14.142
Overid test (p-value)		0.638	0.730	0.006	0.000
Overid. restrictions		1.000	1.000	3.000	5.000

Robust-clustered standard errors in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Additional controls: Year fixed-effects, Firm FE.
Additional characteristics: Air, Doors, HP/WT, DPM, Size, Weight.

terms the number of doors is also positively correlated demand, although the reduced-form parameter is less precisely estimated. Similarly the number of cars with the same "Air" attribute is positively correlated with demand. The joint hypothesis test that the effect of differentiation on demand is zero is rejected with a p-value less than 1%.

An equivalent approach is to test the over-identification restrictions associated with the combined set of instruments: cost-shifter and differentiation. Columns (4)-(5) implement this test by estimating the model under the null that $\lambda_k = 0$ without controlling for z_{jt}^{dist}. Column (5) uses the same differentiation measures as in Column (3), which leads to four restrictions. Column (6) uses a richer set of differentiation measures (see below). The results of the Sargan-Hansen J-test are reported in the bottom panel.

Both sets of moment restrictions are clearly violated, confirming the strenght of the instruments. Importantly, this violation is not caused by the cost-shifters, since we cannot reject the validity of the exclusion restrictions in columns (1)-(3). Note also that the estimated price coefficient is no longer different from zero; consistent with the idea that the differentiation instruments are correlated with prices and the average willingness-to-pay of consumers under the multinomial Logit specification.

Table 12: Definition on the differentiation variables used in the car application

VARIABLES	Definitions	Specifications			
		(1)	(2)	(3)	(4)
$z_{jt,\hat{p}}^{\text{dist}}$	$\sqrt{\sum_{j'}(\hat{p}_{j't}-\hat{p}_{jt})^2}$	✓	✓	✓	✓
$z_{jt,\text{door}}^{\text{dist}}$	$\sqrt{\sum_{j'}(x_{j't,\text{door}}-x_{jt,\text{door}})^2}$	✓	✓	✓	✓
$z_{jt,\text{air}}^{\text{dist}}$	$\sum_{j'\in\mathcal{J}_t}1(x_{j't,\text{air}}=x_{jt,\text{air}})$		✓	✓	✓
$z_{jt,1}^{\text{cov}}$	$\sum_{j'\in\mathcal{J}_t}(\hat{p}_{j't}-\hat{p}_{jt})(x_{j't,\text{door}}-x_{jt,\text{door}})$			✓	✓
$z_{jt,2}^{\text{cov}}$	$\sum_{j'\in\mathcal{J}_t}(\hat{p}_{j't}-\hat{p}_{jt})(x_{j't,\text{door}}-x_{jt,\text{door}})$			✓	✓
$z_{jt,3}^{\text{cov}}$	$\sum_{j'\in\mathcal{J}_t}(\hat{p}_{j't}-\hat{p}_{jt})^2 1(x_{j't,\text{air}}=x_{jt,\text{air}})$				✓
$z_{jt,4}^{\text{cov}}$	$\sum_{j'\in\mathcal{J}_t}(x_{j't,\text{door}}-x_{jt,\text{door}})^2 1(x_{j't,\text{air}}=x_{jt,\text{air}})$				✓
$z_{jt,1}^{\text{int}}$	$\hat{p}_{jt}\times\left(\sum_{j'\in\mathcal{J}_t}(\hat{p}_{j't}-\hat{p}_{jt})\right)$			✓	✓
$z_{jt,2}^{\text{int}}$	$x_{jt,\text{door}}\times\left(\sum_{j'\in\mathcal{J}_t})x_{j't,\text{door}}-x_{jt,\text{door}}\right)$			✓	✓
Number of Diff. IV		2	3	7	9

The GMM results from the mixed-logit model are presented in Table 11. The first column reproduces the results obtained from non-linear least-square. As discussed above, those results are biased, but the precision of the parameters is not affected by the choice of instruments. The comparison with GMM is a useful indicator of validity and relevance of the instruments.

Table 12 formally defines the differentiation measures used as instruments in each specification. The first column uses the Euclidian distance. The second add the number of products with the same Air attribute. The third and fourth column add additional interactions between product characteristics differences. We augment the set of instruments to identify a random-coefficient on the intercept. We use the sum of product characteristics interacted with each product' own attribute as instruments (i.e. $\hat{p}_{jt}$, doors). This captures the fact that in a quality ladder model, low quality cars are closer substitutes with the outside option than higher quality cars. We also incorporate the interaction of product characteristics differences to impose additional restrictions. Table 14 in the Appendix confirms the strenght of these instruments using a series of Sanderson and Windmeijer (2016)'s conditional F-tests associated with each of the five endogenous variables of the model. See Appendix C for more details.

The random coefficient on door and prices are the most precisely estimated coefficients. Note that the overall price coefficient is the sum of β_p and λ_p/y_i. This

coefficient is negative across all consumer types, except households above the 99th income percentile. In 1990, the product-level price elasticity ranges from -0.3 to -3, with a median equal to -1.2. The fact that a large number of products exhibit inelastic residual demand suggest that the model is still mis-specified.

The coefficient on doors and air imply a larger degree of horizontal differentiation. Some consumers strongly prefer two-doors car, while others need the largest number of doors possible (given prices). There is exists a similar amount of heterogeneity across consumers in the taste for air conditioning.

The variance of the random-coefficient on the intercept measures substitution towards the outside good. The point estimate associated with $\lambda_{\text{intercept}}$ is large in magnitude, suggesting that consumers who are in the market for a new car (i.e. high $\eta_{i,\text{intercept}}$) are more likely to substitute to another model if their first choice is not available (rather than not buying a car). The coefficient is less precisely estimated however, given the fact that the panel is fairly short (20 years).

The comparison between column 1 (NLS) and (5) (GMM) confirms that the instruments correct for an attenuation bias in the random-coefficient parameters. In all cases, the asymptotic standard errors from NLS and GMM are comparable, consistent with the idea that the instruments are strong.

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.

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.

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.

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. (1982). The oligopolistic solution concept is identified. *Economic Letters*.

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

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

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

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

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.

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

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

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

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*.

Dube, J.-P., A. Hortacsu, and J. Joo (2020, February). Random-coefficients logit demand estimation with zero-valued market shares. working paper, University of Chicago.

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

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.

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).

Miller, N. and M. C. Weinberg (2017). Understanding the price effects of the millercoors joint venture. *Econometrica 85*(6), 1763–1791.

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. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy 9*(4), 513–548.

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

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

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

Nielson, C. (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.

Salanie, B. and F. A. Wolak (2019, April). Fast, "robust", and approximately correct: Estimating mixed demand systems. NBER working paper 25726.

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

Sullivan, C. (2020, March). The ice cream split: Empirically distinguishing price and product space collusio. Working paper, University of Wisconsin-Madison.

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

Train, K. (2009). *Discrete Choice Methods with Simulation* (Second edition ed.). Cambridge University Press.

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

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

ONLINE APPENDIX - NOT FOR PUBLICATION

A   Proofs and additional derivations

A.1   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*

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

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

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

56

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

$$\mathcal{D}_j(t_0, \ldots, t_J) = \int \mathbf{1}\left[u(t_j, x_j, \theta) \geq u(t_k, x_k, \theta) \ \forall k = 0, \ldots, J, \ k \neq j\right] d\Phi(\theta). \quad (28)$$

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

$$\mathcal{D}_j^{-1}\left(s_0, \ldots, s_J, \boldsymbol{x}^{(2)}\right) = \mathcal{D}^{-1}\left(s_j, \left\{s_k, \boldsymbol{d}_{jk}^{(2)}\right\}_{k \neq j}\right). \quad (29)$$

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 (28) and the symmetry of the idiosyncratic errors (ϵ_{ij}). The first property is

Definition 2. *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 3. *The function* $\mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{-j}, \boldsymbol{x}^{(2)}\right)$ *is* anonymous *if* $\mathcal{D}\left(\boldsymbol{\omega}_j, \boldsymbol{\omega}_{\iota(-j)}, \boldsymbol{x}^{(2)}\right)$ *where* $\iota(\cdot)$ *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 4. *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 (29) and hence Theorem 1.

A.2   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*

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

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

59

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 $\boldsymbol{d}_{jl}$ 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

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

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

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

A.3 Derivation of example with demographic differences

Consider the following single dimension example (Nevo 2001):

$$u_{ijt} = \delta_{jt} + b_{it}x_{jt}^{(2)} + \varepsilon_{ijt}, \quad b_{it} = \lambda_y y_{it} + \eta_i. \tag{30}$$

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:

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

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:

$$
\begin{aligned}
\sigma_{jt}(\boldsymbol{\delta}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}) &= \int \int \frac{\exp\left(\delta_{jt} + \lambda_y y_{it} x_{jt}^{(2)} + \lambda_\eta \eta_i x_{jt}^{(2)}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\delta_{j't} + \lambda_y y_{it} x_{j't}^{(2)} + \lambda_\eta \eta_i x_{j't}^{(2)}\right)} \psi_\eta(d\eta_i; \lambda_\eta) d\Psi_e(y_{it}; \mathrm{m}_t, \mathrm{sd}_t) \\
&= \int \frac{\exp\left(\tilde{\delta}_{jt} + \sum_{k=1}^{K_2} v_{ik} \tilde{x}_{jt,k}^{(2)}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\tilde{\delta}_{j't} + \sum_{k=1}^{K_2} v_{ik} \tilde{x}_{jt,k}^{(2)}\right)} \psi(\boldsymbol{v}_i; \boldsymbol{\lambda}) d\boldsymbol{v}_i \\
&= \sigma_j(\tilde{\boldsymbol{\delta}}_t, \tilde{\boldsymbol{x}}_t^{(2)}; \boldsymbol{\lambda}).
\end{aligned}
\tag{31}
$$

where $\tilde{\boldsymbol{x}}_{jt}^{(2)} = \left\{\mathrm{sd}_t x_{jt}^{(2)}, x_{jt}^{(2)}\right\}$ 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:

$$
\pi_{jt}(\boldsymbol{x}_t; \boldsymbol{\lambda}) = g\left(\boldsymbol{d}_{jt}, \mathrm{sd}_t \cdot \boldsymbol{d}_{jt}^{(2)}; \boldsymbol{\lambda}\right) + c_t(\boldsymbol{\lambda})
\tag{32}
$$

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 instance, the quadratic basis function becomes:

$$
A_j(\boldsymbol{x}_t, \boldsymbol{w}_t) = \begin{cases}
\boldsymbol{w}_{jt}, \mathbf{M}_t \cdot \boldsymbol{x}_{jt}^{(2)} & \text{Own excluded characteristics} \\
\sum_{j' \neq j} \left(d_{jt,j'}^k\right)^2, \quad \forall k & \text{Isolation of product } j \text{ along dimension } k \\
\mathbf{M}_t \cdot \sum_{j' \neq j} \left(d_{jt,j'}^k\right)^2, \quad \forall k & \text{Product isolation} \times \text{demographics} \\
\sum_{j' \neq j} d_{jt,j'}^k \times d_{jt,'}^l, \quad \forall k \neq l & \text{Interaction between dimension } k \text{ and } l \\
\mathbf{M}_t \cdot \sum_{j' \neq j} d_{jt,j'}^k \times d_{jt,'}^l, \quad \forall k \neq l & \text{Characteristics interaction} \times \text{demographics}
\end{cases}
\tag{33}
$$

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

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 13: 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 4 and 4.3 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:

$$\min_{\boldsymbol{\theta}} \quad n g_n(\boldsymbol{\theta}) \boldsymbol{W}_n g_n(\boldsymbol{\theta})^T \tag{34}$$

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:

$$f_j(\boldsymbol{\delta}) = \ln s_{jt} - \ln \sigma_j(\boldsymbol{\delta}_t, \boldsymbol{x}^{(2)}; \boldsymbol{\lambda}) = 0 \quad \forall j = 1, \dots, J. \tag{35}$$

We discretize the distribution of random coefficients to approximate the demand function. To avoid creating an estimation bias, we use the same grid points when simulating and estimating the model. In the multi-dimensional models, we use the quartiles of the standard-normal along each dimension. With a single dimension, this leads to $S = 4$ consumer types. With two dimensions, $S = 16$. Etc. The weights assigned to each consumer type sum to one, and are proportional to the normal density.

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:*

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

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:

$$\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}) = \boldsymbol{Z}\boldsymbol{\pi}\boldsymbol{b} + \boldsymbol{v}^1 \tag{36}$$

$$\boldsymbol{J}(\hat{\boldsymbol{\theta}}) = \boldsymbol{Z}\boldsymbol{\pi} + \boldsymbol{v}^2 \tag{37}$$

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   Gauss-Newton regression and weak IV tests

We analyze the weakness of the instruments by estimating a Gauss-Newton (GN) regression. The GN regression is most often used to minimize GMM or non-linear least-square problem. The solution to the non-linear GMM problem can be obtained as a sequence of linear GMM regressions obtained from a quadratic approximation of the moments. Salanie and Wolak (2019) uses this property to construct a "fast" estimator of the mixed-logit model. The GN regression can also be used to conduct inference on $\hat{\theta}^{gmm}$, since the variance-covariance matrix of $\hat{\theta}^{gmm}$ is the same as linear coefficients from the GN regression. See Davidson and MacKinnon (2001) for a discussion of the

usefulness of "artificial regressions" for testing and inference in econometrics.

Conditional of a vector of parameters ($\theta = \beta, \lambda$), the GN regression is given by linear structural and reduced-form equations:

$$\boldsymbol{\rho}(\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{\theta})\boldsymbol{b} + \boldsymbol{v}^1 \tag{38}$$

$$\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{Z}\boldsymbol{\pi} + \boldsymbol{v}^2 \tag{39}$$

where $\boldsymbol{J}(\hat{\boldsymbol{\theta}})$ is a $n \times |\boldsymbol{\theta}|$ matrix containing the slopes of the inverse demand with respect to each of the non-linear parameters. That, is $J_{jt,k}(\boldsymbol{\theta}) = \partial\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\lambda}^k\right)/\partial\lambda_k$ if $\theta_k = \lambda_k$ (non-linear parameter), and $J_{jt,k}(\boldsymbol{\theta}) = -x_{jt,k}$ if $\theta_k = \beta_k$ (linear parameter). This system of equations contains $|\boldsymbol{\lambda}|+1$ endogenous variables: the Jacobian associated with each of the random coefficient parameter and price. The solution to the non-linear GMM problem is defined as $\hat{b}(\theta^{gmm}) = 0$, since $\hat{b}$ is the score of the GMM problem. We use the GN regression evaluated at $\hat{\theta}^{gmm}$ to conduct inference on the parameters, and test the rank condition of the non-linear GMM problem.

Table 14 reports the Sanderson and Windmeijer (2016)'s conditional F-tests associated with each of the five endogenous variables of the model. We construct these weak IV tests using the Jacobian function of the residual evaluated at the GMM estimates.[30] The statistics are all above standard levels for first-stage F-tests, consistent with the idea that the instruments generate enough independent variation in the moments to identify all five parameters.

Note that we do not report the p-values associated with each test. This is because the F-tests are evaluated at the estimated parameters, rather than at the true.[31] This implies that standard statistical significance tables are not necessarily applicable. Our simulations results however show that statistics evaluated at θ^0 and $\hat{\theta}$ are highly correlated, even in the presence of weak instruments. We therefore interpret the magnitude of the tests as measure of strength.

Below is a sample STATA code to estimate the Gauss-Newton regression to conduct inference and tests for weak instruments. The code uses two input files. "Results_gnr_$spec.dta" is a spreadhseet containing the product characteristics, instruments, and Jacobian functions. "Results_wmatrix_$spec.dta" is a square matrix containing the (inverse) weighing matrix used for the minimization of the GMM problem.

---

[30]See Appendix C for more details.

[31]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.

Table 14: First-stage Weak IV tests using the car data

Specifications	Test statistics	λ_{price}	λ_{doors}	λ_{air}	$\lambda_{\mathrm{intercept}}$	β_{price}
					Variables	
(1)	SW - F	19.32	111.88			34.35
(2)	SW - F	12.04	77.98	81.55		20.70
(3)	SW - F	20.50	66.49		90.19	185.68
(4)	SW - F	14.73	61.89	41.22	36.44	116.12

The code uses the user-written ado file "ivreg2" which estimates linear GMM model, and runs a battery of specification tests. The Jacobian functions of the inverse demand are calculated analytically using the implicit function theorem (see Nevo (2000)).

```
#delimit;


use "Results_gnr_$spec.dta", clear;


reshape long diffiv_, i(model year) j(iv);
sum iv;
global nb_diffiv=r(max);
global diffiv;
forvalues i=0/$nb_diffiv {;
global diffiv $diffiv diffiv_'i';
};
global iv $priceiv $diffiv;


/* Weighting matrix */
use "Results_wmatrix_$spec",clear; /* Load weighting matrix */


mkmat $x Intercept $iv , matrix(A); /* Create matrix type */
mat rownames A=$x _cons $iv ;
mat colnames A=$x _cons $iv;
mat W=invsym(A);


/* Load residual, characteristics, jacobian and IVs */
use "Results_gnr_$spec", clear;
```

```
format * %9.6g;

/* GN IV regression with weighting matrix W */
ivreg2 xi $x ($endovar =$iv), r wmatrix(W) cluster(model) ffirst;
mat V=diag(vecdiag(e(V)));
mat se_$spec=vecdiag(cholesky(V))';
mat colnames se_$spec=se_$spec;
mat spec_test_$spec=(e(idstat),e(idp),e(j),e(jp),e(jdf))';
mat rownames spec_test_$spec= KP_test KP_pv J_test J_pv J_df;


/* Save weak IV tests */
clear;
mat F=e(first);
svmat2 F, rnames(stats) names(col);
keep if stats=="SWF" | stats=="SWFp";
gen spec="$spec";
list;
mat list F;
save results/WeakIV_test_$spec, replace;

use results/Car_demand_estimates_$spec, clear;
rename Var1 est_$spec;
mkmat est;
mat rownames est_$spec= $endovar $x _cons;


/* Save GMM results to spreadsheet */
clear;
svmat2 est_$spec, rnames(param);
svmat2 se_$spec;
local nparam=_N;
local N=_N+5;
set obs `N';
```

```
svmat2 spec_test_$spec, rnames(stats);
local i=1;
local row='nparam'+1;
foreach var in KP_lm KP_pv J_test J_pv J_df {;
replace est_$spec=spec_test_$spec1['i'] if _n=='row';
replace param=stats['i'] if _n=='row';
local row='row'+1;
local i='i'+1;
};
drop stats;
drop spec_test_$spec1;
save "results/GMM_results_$spec", replace;
```