

# Included and Excluded Instruments in Structural Estimation

Isaiah Andrews, *Harvard University and NBER\**

Nano Barahona, *UC Berkeley*

Matthew Gentzkow, *Stanford University and NBER*

Ashesh Rambachan, *Harvard University*

Jesse M. Shapiro, *Harvard University and NBER*

April 2022

## Abstract

We consider the choice of instrumental variables when a researcher's structural model may be misspecified. We contrast *included instruments*, which have a direct causal effect on the outcome holding constant the endogenous variable of interest, with *excluded instruments*, which do not. We show conditions under which the researcher's estimand maintains an interpretation in terms of causal effects of the endogenous variable under excluded instruments but not under included instruments. We apply our framework to estimation of a linear instrumental variables model, and of differentiated goods demand models under price endogeneity. We show that the distinction between included and excluded instruments is quantitatively important in simulations based on an application. We extend our results to a dynamic setting by studying estimation of production function parameters under input endogeneity.

---

\*E-mail: iandrews@fas.harvard.edu, nanobk@berkeley.edu, gentzkow@stanford.edu, asheshr@g.harvard.edu, jesse\_shapiro@fas.harvard.edu. We acknowledge funding from the National Science Foundation (SES-1949047, SES-1949066, DGE1745303), the Sloan Foundation, the Brown University Population Studies and Training Center, the Stanford Institute for Economic Policy Research, and the Eastman Professorship at Brown University. We thank our dedicated research assistants for their contributions to this project, Nathan Miller and Matthew Weinberg for help with their code and data, and Tim Armstrong and Larry Katz for helpful comments. All estimates and analyses in this paper based on Information Resources Inc. data are by the authors and not by Information Resources Inc.

# 1 Introduction

A researcher observes data on an outcome variable  $Y_i$ , an endogenous variable of interest  $D_i$ , and exogenous variables  $(X_i, Z_i)$  for each of a set of units  $i$ . The vector  $X_i$  consists of *included* variables that may have a causal effect on  $Y_i$ ; the vector  $Z_i$  consists of *excluded* variables that do not. The researcher, who is interested in measuring the causal effect of the endogenous variable  $D_i$  on  $Y_i$ , models the outcome as  $Y_i = Y^*(D_i, X_i, \xi_i, \varepsilon_i; \theta)$  where  $\theta$  is a finite-dimensional vector of parameters,  $\xi_i$  is a  $J$ -dimensional structural residual independent of the exogenous variables  $(X_i, Z_i)$ ,  $\varepsilon_i$  is a vector of additional structural errors of unrestricted dimension, and  $Y^*(\cdot)$  is a known function.

The researcher's model implies that  $\xi_i = R^*(Y_i, D_i, X_i; \theta)$  where  $R^*(\cdot)$  is a known function. The researcher therefore estimates the parameter vector  $\theta$  using a moment condition formed as the product of  $R^*(Y_i, D_i, X_i; \theta)$  and instrument functions that depend on  $(X_i, Z_i)$

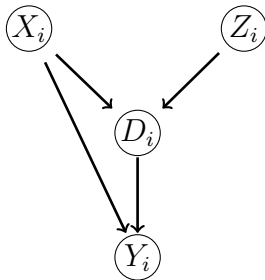
$$E \left[ \sum_{j=1}^J R_j^*(Y_i, D_i, X_i; \theta) f_j(X_i, Z_i) \right] = 0 \quad (1)$$

for known  $f_j(\cdot)$ . This setup nests linear instrumental variables as well as common estimators of many nonlinear models with endogeneity.

Discussions of the choice of instruments  $f(X_i, Z_i)$  often turn on the strength of the relationship of  $f(X_i, Z_i)$  to  $D_i$  and the plausibility of the orthogonality of  $f(X_i, Z_i)$  and  $\xi_i$ . We focus instead on the sensitivity of the researcher's conclusions to misspecification of the form of  $Y^*(\cdot)$ . We show that this consideration alone can provide a reason to prefer some instruments over others, even when all instruments are strong and as-good-as-randomly assigned.

To capture the possibility of misspecification, we embed the researcher's model in a broader class of models in which  $Y_i = Y_i(D_i, X_i)$ , for  $Y_i(\cdot)$  a random potential outcome function. We assume that  $Y_i(\cdot)$  is independent of  $(X_i, Z_i)$ , thus maintaining the analogue of the independence of the structural residual in the researcher's model. We also assume that  $X_i$  can directly influence the outcome  $Y_i$ , whereas  $Z_i$  can influence the outcome  $Y_i$  only via  $D_i$ . In this sense, we maintain the same underlying causal structure as in the researcher's model (see Figure 1) but allow for arbitrary, global misspecification of the researcher's conjectured functional form  $Y^*(\cdot)$ . We refer to functions of  $X_i$  as *included instruments* and to functions of  $Z_i$  as *excluded instruments*. Section 2 provides this formal setup.

Figure 1: Directed acyclic graph summarizing the causal relationships in the model



In a linear instrumental variables setting where  $Y_i^*(D_i, X_i, \xi_i; \theta) = \alpha D_i + X_i' \beta + \xi_i$ , for  $\alpha, \beta$  unknown parameters, we might have that, say, the outcome  $Y_i$  is the log of earnings for individual  $i$ , the excluded variable  $Z_i$  is a vector of quarter-of-birth indicators, and the included variables  $X_i$  include age and its square (Angrist and Krueger 1991). Excluded instruments would include functions of quarter of birth indicators; included instruments would include functions of age not captured in  $X_i$  that are nevertheless correlated with the residual variation in  $D_i$ . Included instruments are valid under the assumed model  $Y_i^*(\cdot)$  because the model implies that the term  $X_i' \beta$  fully captures the causal effect of age on log earnings.

Both excluded and included instruments are used frequently in structural research. One leading example is differentiated goods demand models (e.g., Berry et al. 1995), where  $i$  is a market,  $Y_i$  is a vector of market shares,  $D_i$  is a vector of prices,  $X_i$  is a vector of product characteristics, and  $Z_i$  might include cost shifters that affect firms' optimal prices but do not otherwise affect demand (e.g., Nevo 2001). Because a firm's price is affected by the characteristics of other firms in the market, functions of those characteristics may be available as included instruments, and are sometimes referred to as "BLP instruments," following their introduction by Berry et al. (1995). A second leading example is production function estimation (e.g., Akerberg et al 2007, Section 2), where  $i$  is a firm,  $Y_i$  is a sequence of physical outputs,  $D_i$  is a sequence of inputs,  $X_i$  is a sequence of state variables including past input choices, and  $Z_i$  might include input prices that affect firms' optimal input choices but do not otherwise affect output (e.g., de Roux et al. 2021). Because a firm's current input is related to its past input choices, functions of past inputs may be available as included instruments (Blundell and Bond 1998, 2000; Akerberg et al 2015).<sup>1</sup>

Section 3 provides a key proposition characterizing the causal interpretation of the researcher's estimand  $\tilde{\theta}$  that solves (1) under the potential outcome model. The proposition shows that, in

<sup>1</sup>Other examples include treating the product of two included variables as an instrument (e.g., Gentzkow and Shapiro 2008), as well as some of the approaches discussed in Lewbel (2019, section 3.7).

population, the researcher's estimand zeros out a particular weighted average of causal effects of  $X_i$  and  $Z_i$  on the residual function  $R^*(Y_i, D_i, X_i; \theta)$ . This characterization holds under a potential outcome model that allows for arbitrary (global) misspecification of the researcher's functional form assumptions  $Y^*(\cdot)$  and any choice of the residual function. The economic interpretation of the results is clearer with additional structure on the researcher's residual function. Accordingly, we provide several results that further characterize the researcher's estimand for the prominent special case in which the residual function is linear in the endogenous variable of interest  $D_i$ . In the rest of the paper, we apply the results from this special case to analyze several leading structural models used in applied research.

Section 4 applies the characterization in Section 3 to the estimation of a linear instrumental variables (IV) model in order to build intuition. We show that the usual linear IV estimand has an interpretation in terms of causal effects of  $D_i$  on  $Y_i$  when using excluded instruments, but not, in general, when using included instruments. The analysis of this application connects our results to those of Angrist et al. (2000) for linear simultaneous equation models.

Section 5 applies the characterization in Section 3 to the estimation of differentiated goods demand models with price endogeneity. When the researcher estimates a random-coefficients logit demand model (Berry et al. 1995) using excluded instruments, the model-implied, *ceteris paribus* effects of product prices on market shares, evaluated at the researcher's estimand, match the true causal effects on average. In the special case where the researcher does not allow random coefficients and therefore estimates a multinomial logit model, the researcher's price coefficient recovers a particular weighted average of semi-elasticities of market shares with respect to product prices. Although the weights in these averages need not be positive, these results provide a sense in which excluded instruments guarantee a causal interpretation of the researcher's estimand even under misspecification. In contrast, when using included instruments such as functions of product characteristics, this causal interpretation is lost. To the best of our knowledge, these results provide the first nonparametric causal interpretation of the estimands of differentiated goods demand models.

Section 5 also presents an application to the estimation of the demand for beer in Miller and Weinberg (2017). We calibrate simulations tightly to Miller and Weinberg's (2017) estimated model and mimic the behavior of a researcher who uses instruments to address price endogeneity. When the researcher uses excluded instruments, the researcher's estimates of the average own-price elasticity exhibit relatively small median bias even when the researcher's model is misspecified. By contrast, when the researcher uses included instruments, the researcher's estimates of the average

own-price elasticity can be severely median-biased under misspecification, so much so that in some cases a researcher concerned about bias would be better off ignoring the need for instruments altogether and (erroneously) treating the price as exogenous.

Finally, Section 6 extends our analysis to dynamic settings by studying the estimation of production function models with input endogeneity. We characterize the researcher’s estimand under contemporaneous, excluded instruments when the researcher estimates a Cobb-Douglas production model with a serially dependent productivity shock (Ackerberg et al. 2015). We show that the researcher’s coefficient on the contemporaneous static input recovers a particular (non-convex) weighted average of true output elasticities defined under the potential outcomes model. When using included instruments, such as past values of inputs, this interpretation is lost.

Taking our theoretical and numerical findings together, we advocate prioritizing excluded over included instruments. In cases where excluded instruments are not available, and the researcher lacks a strong justification for the chosen form of  $Y_i^*(\cdot)$ , we advocate providing additional sensitivity analysis with respect to other *a priori* reasonable choices of  $Y_i^*(\cdot)$ .

A large literature following Imbens and Angrist (1994) and Angrist et al. (1996) studies the interpretation of instrumental variables estimators under potential model misspecification. Within this literature our work is closest to that of Angrist et al. (2000), who study the nonparametric interpretation of estimands in linear simultaneous equations models when excluded instruments are used.<sup>2</sup> Our contributions are to consider settings in which the outcome variable is potentially vector-valued and the researcher’s model is potentially nonlinear, and to characterize estimands under both excluded and included instruments. Our results are applicable to important economic contexts in which nonlinear structural models are estimated using instruments, for which (to our knowledge) a similar characterization of estimands was not previously available.

Recent work has studied issues of nonparametric identification in the settings we consider. See, for example, Berry and Haile (2014, 2016) in the context of differentiated goods demand models, and Gandhi et al. (2020) in the context of production models. Instead of studying nonparametric identification of objects of economic interest, we study the nonparametric interpretation of the

---

<sup>2</sup>Angrist (2001) studies IV estimands in limited dependent variable settings, and characterizes a nonlinear estimand in terms of causal effects. Kolesár (2013) and Andrews (2019) compare the estimands of different IV estimators in linear models. Chalak (2017) studies the interpretation of IV estimands in ordered discrete choice settings under violations of monotonicity. Mogstad et al. (2018) discuss the interpretation of linear IV estimands in terms of marginal treatment effect functions. Kline and Walters (2019) show that many nonlinear and linear models deliver numerically equivalent estimates for local average treatment effects and average potential outcomes among certain subgroups. Blandhol et al. (2022) and Słoczyński (2022) provide conditions under which two stage least squares specifications that parametrically control for covariates estimate a weighted average of causal effects.

researcher’s estimator when the researcher specifies and estimates a particular model. These two exercises are distinct but related, and we highlight some connections in Sections 5 and 6.

Our work is also related to large literatures on the optimal choice of instruments or moment conditions. One part of that literature (e.g., Hansen 1982; Chamberlain 1987; Newey 1990) studies efficiency considerations under the assumption of correct specification.<sup>3</sup> Another part of the literature discusses the economic validity of the orthogonality conditions underlying various included or excluded instruments in particular structural models.<sup>4</sup> A third part of this literature (e.g., Kitamura et al. 2013; Armstrong and Kolesár 2021; Bonhomme and Weidner forthcoming) seeks procedures with optimality properties under certain forms of potential misspecification.<sup>5</sup> Our analysis highlights a broad distinction between moment conditions formed using excluded versus included instruments under rich forms of possible misspecification. The distinction we draw between excluded and included instruments is related to Akerberg et al.’s (2011) suggestion to learn the effect on an outcome of one endogenous variable in the presence of a second endogenous variable by employing instruments that are orthogonal to the second variable.

## 2 Setup

The sample consists of  $n$  observations  $(Y_i, D_i, X_i, Z_i) \in \mathcal{Y} \times \mathcal{D} \times \mathcal{X} \times \mathcal{Z}$  drawn i.i.d. from an unknown distribution  $G$ . To model causal effects we adopt a potential outcomes framework, where the distribution of the potential outcome and potential endogenous variable functions  $Y_i(\cdot)$  and  $D_i(\cdot)$  is again governed by  $G$ , and the realized values are  $Y_i = Y_i(D_i, X_i, Z_i)$  and  $D_i = D_i(X_i, Z_i)$ .

**Assumption 1.** *(General model) Under the true data-generating process, the following hold:*

- (a) *(exclusion) For all  $Y_i(\cdot)$ ,  $Y_i(d, x, z) = Y_i(d, x, z') = Y_i(d, x)$  for all  $d \in \mathcal{D}$ ,  $x \in \mathcal{X}$ , and  $z, z' \in \mathcal{Z}$ .*

---

<sup>3</sup>Reynaert and Verboven (2014) study these issues in the context of Berry et al.’s (1995) estimator. A related literature (e.g., Rossi 2014; Armstrong 2016; Gandhi and Houde 2020) considers the possibility of weak instruments in this setting. Gandhi and Houde (2020) recommend using carefully chosen functions of included variables as instruments in order to improve instrument strength.

<sup>4</sup>In the context of differentiated goods demand models, see, for example, Bresnahan (1996), Nevo (2004), Rossi (2014), and Petrin and Seo (2019).

<sup>5</sup>Analytically, our approach differs from much of this literature in that we consider misspecification that is nonlocal, in the sense that the degree of misspecification remains fixed as the sample grows large. Hall and Inoue (2003) characterize the asymptotic distribution of GMM estimators under nonlocal misspecification.

(b) (*random assignment*)  $(Y_i(\cdot), D_i(\cdot)) \perp\!\!\!\perp X_i \perp\!\!\!\perp Z_i$ .

Assumption 1(a) states that the excluded variables  $Z_i$  do not causally affect the outcome  $Y_i$  except through the endogenous variable  $D_i$ , as in Figure 1. Assumption 1(b) states that  $X_i$  and  $Z_i$  are independent of the unobservable determinants of the outcome and endogenous variable,  $Y_i(\cdot)$  and  $D_i(\cdot)$ , and of each other. Thus, we will interpret both  $X_i$  and  $Z_i$  as exogenous variables, where the two are distinguished by the assumption that only  $X_i$  may have a direct causal impact on the outcome  $Y_i$ . Assumption 1(b) is stronger than requiring  $Z_i$  to be independent of  $(Y_i(\cdot), D_i(\cdot))$  conditional on  $X_i$ , as in, e.g., Angrist et al. (2000). Imposing exogeneity of both  $X_i$  and  $Z_i$  helps clarify the distinction we draw between included and excluded instruments. Our characterizations of estimands under excluded instruments remain unchanged, however, if we drop the exogeneity of  $X_i$  by weakening Assumption 1(b) to  $(Y_i(\cdot), D_i(\cdot), X_i) \perp\!\!\!\perp Z_i$ .<sup>6</sup>

## 2.1 Researcher's Model

The researcher models the outcome, assuming that  $Y_i(d, x) = Y^*(d, x, \xi_i, \varepsilon_i; \theta_0)$ , for  $\theta_0 \in \mathbb{R}^P$  an unknown parameter,  $Y^*(\cdot)$  a known function, and  $\xi_i, \varepsilon_i$  structural residuals. The residual  $\xi_i$  is  $J$ -dimensional, which in many applications is also the dimension of  $Y_i$ . If we knew  $\theta_0$ , the residual could be recovered by taking an appropriate transformation of the data; that is,  $\xi_i = R^*(Y_i, D_i, X_i; \theta_0)$  for  $R^*(\cdot)$  a known function. By contrast, the dimension of  $\varepsilon_i$  is unrestricted, and we may not be able to recover  $\varepsilon_i$  from the observed data (for instance because the number of shocks exceeds the number of observables).

**Assumption 2.** (*Researcher's model*) Under the researcher's model, the following hold:

(a) (*outcome model*)  $Y_i(d, x) = Y^*(d, x, \xi_i, \varepsilon_i; \theta_0)$  and  $\xi_i = R^*(Y_i(d, x), d, x; \theta_0)$ .

(b) (*residual independence*)  $(\xi_i, \varepsilon_i) \perp\!\!\!\perp X_i \perp\!\!\!\perp Z_i$ .

(c) (*mean-zero residual*)  $E[\xi_i] = 0$ .

Assumption 2(a) imposes the researcher's outcome model and requires that  $\xi_i$  could be recovered if  $\theta_0$  were known. Assumption 2(b) imposes that the residuals  $(\xi_i, \varepsilon_i)$  are independent of  $(X_i, Z_i)$ . Finally, Assumption 2(c) requires that the residual  $\xi_i$  has mean zero.

<sup>6</sup>Both Assumption 1(b) and  $(Y_i(\cdot), D_i(\cdot), X_i) \perp\!\!\!\perp Z_i$  are stronger than conditional independence given  $X_i$ , i.e.,  $(Y_i(\cdot), D_i(\cdot)) \perp\!\!\!\perp Z_i | X_i$ . Under conditional independence, analogs of our results for excluded instruments can be obtained for estimators that account nonparametrically for the dependence between  $X_i$  and  $Z_i$ , e.g., by weighting observations inversely by  $p(Z_i | X_i)$ , for  $p(z|x) = Pr\{Z_i = z | X_i = x\}$ .

Assumptions 2(b) and 2(c) together imply that  $\xi_i$  has conditional mean zero given  $(X_i, Z_i)$ , i.e.,  $E[\xi_i | X_i, Z_i] = 0$ , a widely-imposed assumption in structural estimation. Likewise, Assumptions 2(a) and 2(b) together imply that  $Y_i(\cdot) \perp\!\!\!\perp (X_i, Z_i)$ , consistent with Assumption 1(b). Together, Assumptions 1 and 2 allow us to characterize estimands under both the researcher's model and under a general potential outcomes model that nests the researcher's model.

## 2.2 Researcher's Estimator

Assumption 2 implies that the product of  $R^*(Y_i, D_i, X_i; \theta_0)$  with any function of  $(X_i, Z_i)$  has mean zero. Hence, the researcher can estimate  $\theta$  by GMM, constructing moments as the product of residual function  $R^*(Y_i, D_i, X_i; \theta)$  with instrument functions depending on  $(X_i, Z_i)$ . We assume that, whether or not the model is correct, the resulting estimator  $\hat{\theta}$  is consistent for some solution to a just-identified system of moments.

**Assumption 3.** (*Just-identification*) For  $R_j^*(Y_i, D_i, X_i; \theta)$  the  $j$ th element of  $R^*(Y_i, D_i, X_i; \theta)$ , the estimator  $\hat{\theta}$  converges in probability to a value  $\tilde{\theta}$  that solves

$$E \left[ \sum_{j=1}^J R_j^*(Y_i, D_i, X_i; \theta) f_j(X_i, Z_i) \right] = 0, \quad (2)$$

where  $f_j(X_i, Z_i) \in \mathbb{R}^P$  for each  $j$ .

Under regularity conditions (see Newey and McFadden 1994, Theorem 2.6), Assumption 3 holds if the researcher chooses  $\hat{\theta}$  to solve

$$\frac{1}{n} \sum_i \sum_j R_j^*(Y_i, D_i, X_i; \theta) f_j(X_i, Z_i) = 0.$$

Assumption 3 is more general than this and, for  $f_j$  defined appropriately, also holds when the researcher estimates an over-identified GMM model.<sup>7</sup> Crucially, the conditions for consistency of  $\hat{\theta}$  for  $\tilde{\theta}$  do not require that the researcher's model is correct in the sense of Assumption 2.

<sup>7</sup>Specifically, with basic instrument vectors  $f_j^*(X_i, Z_i)$  of dimension strictly greater than  $P = \dim(\theta)$ , sample moments  $\hat{m}(\theta) = \frac{1}{n} \sum_i \sum_j R_j^*(Y_i, D_i, X_i; \theta) f_j^*(X_i, Z_i)$ , and a weighting matrix  $\hat{\Omega}$ , the GMM estimator  $\hat{\theta}$  minimizes  $\hat{m}(\theta)' \hat{\Omega} \hat{m}(\theta)$ . Under regularity conditions (see Newey and McFadden 1994, Theorem 2.1),  $\hat{\theta}$  will be consistent for the value  $\tilde{\theta}$  which solves  $\min_{\theta} m(\theta)' \Omega m(\theta)$  for  $m(\theta) = E \left[ \sum_{j=1}^J R_j^*(Y_i, D_i, X_i; \theta) f_j^*(X_i, Z_i) \right]$  and  $\Omega$  the probability limit of  $\hat{\Omega}$ . Critically, these results apply even if the model is misspecified, in the sense that  $m(\theta) \neq 0$  for all  $\theta$ . Provided  $\tilde{\theta}$  is an interior solution, Assumption 3 then holds for  $f_j(X_i, Z_i) = \frac{\partial}{\partial \theta} m(\tilde{\theta})' \Omega f_j^*(X_i, Z_i)$ , where the Jacobian  $\frac{\partial}{\partial \theta} \tilde{m}(\tilde{\theta})$ , and thus the functions  $f_j$ , can be consistently estimated.



### 3 Characterization of Researcher's Estimand

If Assumption 2 holds and the solution to (2) is unique, then  $\tilde{\theta} = \theta_0$  and  $\hat{\theta}$  is consistent for  $\theta_0$ .

**Assumption 4.**  $\tilde{\theta}$  is the only value of  $\theta$  that solves (2).

**Lemma 1.** Under Assumptions 2, 3, and 4,  $\hat{\theta} \rightarrow_p \theta_0 = \tilde{\theta}$ .

All proofs, including for Lemma 1, are given in Appendix B.

While  $\hat{\theta}$  is consistent for  $\theta_0$  when the researcher's model is correct, it is less clear how to interpret this estimator under misspecification. To answer this question, we relate the researcher's estimand  $\tilde{\theta}$  to the causal effect of  $D_i$  under the potential outcomes model. To do so, we require one further assumption, namely that some linear transformation  $v' f_j$  of the instrument vectors  $f_j$  has mean zero for all  $j$ .

**Assumption 5. (Demeaned instruments)** There exists nonzero  $v \in \mathbb{R}^P$  such that for  $f_j^v(X_i, Z_i) = v' f_j(X_i, Z_i)$ ,  $E[f_j^v(X_i, Z_i)] = 0$  for all  $j$ .

Assumption 5 holds if at least one element of the instrument vector  $f_j(X_i, Z_i)$  has been demeaned for all  $j$ . It also holds whenever  $P > J$ , for instance when  $\theta$  includes at least one  $j$ -specific parameter for each  $j$  (such as a brand fixed effect) and at least one common parameter (such as a price coefficient). Assumption 5 is therefore satisfied automatically in many applications. Even when this is not the case, researchers may ensure that Assumption 5 holds by de-meaning one or more elements of their instrument vector.

Assumption 5 allows us to re-express expectations involving  $f_j^v(X_i, Z_i)$  in terms of differences in weighted averages. Specifically, define probability distributions  $H_+(\cdot|j)$  and  $H_-(\cdot|j)$  on  $\mathcal{X} \times \mathcal{Z}$  by

$$H_+(\mathcal{S}|j) = \int_{\mathcal{S}} \frac{\max\{f_j^v(x, z), 0\}}{\frac{1}{2}E\left[\left|f_j^v(X_i, Z_i)\right|\right]} dG(x, z), \quad H_-(\mathcal{S}|j) = \int_{\mathcal{S}} \frac{\max\{-f_j^v(x, z), 0\}}{\frac{1}{2}E\left[\left|f_j^v(X_i, Z_i)\right|\right]} dG(x, z) \quad (3)$$

for all sets  $\mathcal{S}$ . Further define

$$h(j) = \frac{E\left[\left|f_j^v(X_i, Z_i)\right|\right]}{\sum_{j'} E\left[\left|f_{j'}^v(X_i, Z_i)\right|\right]}.$$

For any function  $b(X_i, Z_i)$ ,

$$\sum_j E\left[b(X_i, Z_i) f_j^v(X_i, Z_i)\right] \propto \sum_j \int b(x, z) dH_+(x, z|j) h(j) - \sum_j \int b(x, z) dH_-(x, z|j) h(j)$$

by construction.<sup>8</sup>

Combined with Assumption 1, this allows us to re-cast the moment condition (2) in terms of causal effects.

**Definition 1.** (Causal effects) The **causal effect** of changing  $(d_-, x_-)$  to  $(d_+, x_+)$  on  $B(Y_i, D_i, X_i)$  is

$$\tau_i^B(d_+, x_+, d_-, x_-) \equiv B(Y_i(d_+, x_+), d_+, x_+) - B(Y_i(d_-, x_-), d_-, x_-).$$

The causal effect of changing  $d_-$  to  $d_+$  on  $B(Y_i, D_i, X_i)$ , holding  $x$  fixed, is

$$\tau_i^B(d_+, d_-, x) \equiv B(Y_i(d_+, x), d_+, x) - B(Y_i(d_-, x), d_-, x).$$

**Proposition 1.** Under Assumptions 1, 3, and 5,

$$\sum_j \int \int E \left[ \tau_i^{R_j^*} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j) = 0 \quad (4)$$

for  $\tau_i^{R_j^*} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right)$  the causal effect of changing  $(x_-, z_-)$  to  $(x_+, z_+)$  on  $R_j^*(Y_i, D_i, X_i; \tilde{\theta})$ .

Hence, at  $\theta = \tilde{\theta}$  the  $h(j)$ -weighted average causal effect on the residual  $R_j^*(Y_i, D_i, X_i; \theta)$  of changing the distribution of  $(X_i, Z_i)$  from  $H_-(x, z|j)$  to  $H_+(x, z|j)$  is zero. Under Assumptions 2 and 4,

$$\tau_i^{R_j^*} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) = \xi_i - \xi_i = 0,$$

so (4) holds trivially under the researcher's model. The value of Proposition 1 thus lies in showing that (4) continues to hold even under the potential outcome model. This gives us a route to a causal interpretation of  $\tilde{\theta}$  even under global misspecification of the researcher's conjectured functional form  $Y^*(\cdot)$ .

The distributions  $H_+(\cdot|j)$  and  $H_-(\cdot|j)$  and weights  $h(j)$  depend on the linear combination of instruments  $f_j^v(X_i, Z_i)$  considered. Hence, if there are linearly independent vectors  $v$  and  $\tilde{v}$  that both satisfy Assumption 5, Lemma 1 implies multiple weighted average interpretations, each in terms of different weights. Appendix A.2 characterizes the full set of weights for which (4) holds, and shows that, under an additional condition, all weights in this set correspond to some  $v$  satisfying Assumption 5.

---

<sup>8</sup>Assumption 5 is crucial for this argument, as otherwise  $H_+$  and  $H_-$  integrate to different values, so at least one cannot be interpreted as a probability distribution. See Appendix A.1 for further discussion.

Proposition 1 relates  $\tilde{\theta}$  to the causal effect of simultaneously changing both  $D_i$  and  $X_i$ , rather than the *ceteris paribus* effect of changing only  $D_i$ . The direct effects of  $X_i$  drop out when the researcher uses excluded instruments.

**Definition 2.** (Excluded instruments) The researcher uses **excluded instruments** if, for some  $v$  satisfying Assumption 5, and some  $\tilde{f}_j$ ,  $f_j^v(x, z) = \tilde{f}_j(z)$  for all  $(x, z)$  and all  $j$ . Otherwise, the researcher uses **included instruments**.

Thus, the researcher uses excluded instruments if some linear combination of their instrument functions both (i) satisfies Assumption 5 and (ii) depends only on  $z$ . The simplest case is where the researcher instruments exclusively with functions of  $z$ , i.e.,  $f_j(x, z) = f_j(z)$  for all  $j$ . It is often helpful to include some functions of  $X_i$  among the instruments however, e.g., to estimate parameters relating  $X_i$  to  $Y_i$ . Our definition of excluded instruments allows for this possibility.

When the researcher uses excluded instruments,  $\tilde{\theta}$  can be interpreted in terms of the causal effect of  $D_i$  alone. To state this result, let  $H_+(z|j)$  and  $H_-(z|j)$  be distributions on  $\mathcal{Z}$  defined as in (3), replacing  $f_j^v(x, z)$  by  $\tilde{f}_j(z)$ .

**Corollary 1.** *Under Assumptions 1, 3, and 5, if the researcher uses excluded instruments, then*

$$\sum_j \int \int E \left[ \tau_i^{R_j^*} \left( D_i(X_i, z_+), D_i(X_i, z_-), X_i; \tilde{\theta} \right) \right] dH_+(z_+|j) dH_-(z_-|j) h(j) = 0 \quad (5)$$

for  $\tau_i^{R_j^*}(d_+, d_-, x; \tilde{\theta})$  the causal effect of changing  $d_-$  to  $d_+$  on  $R_j^*(Y_i, D_i, X_i; \tilde{\theta})$ .

Hence, when the researcher uses excluded instruments,  $\tilde{\theta}$  sets a particular average causal effect of  $D_i$  on the model-implied residuals equal to zero. Since residuals need not be of direct economic interest, we want to translate (5) into a more economically interpretable form. The extent to which this can be done depends on the structure of the residual function  $R^*$ , and for the rest of the paper we focus on the leading case where  $R^*$  is linear in  $D_i$ .

### 3.1 Special Case: Linear Residual

**Definition 3.** The researcher uses a **linear residual** if  $D_i$  is  $J$ -dimensional and

$$R_j^*(Y_i, D_i, X_i; \theta) = \delta_j(Y_i, X_i; \theta) - \alpha D_{i,j}$$

for  $\delta(\cdot)$  a known  $J$ -dimensional function and  $\alpha$  an element of  $\theta$ .

All of the applications we discuss have linear residuals. For linear residuals, we obtain an explicit expression for  $\tilde{\alpha}$ , the element of  $\tilde{\theta}$  corresponding to  $\alpha$ , as a ratio of weighted average causal effects.

**Corollary 2.** *Under Assumptions 1, 3, 5, if the researcher uses a linear residual and*

*$\sum_j E[D_{i,j} f_j^v(X_i, Z_i)] \neq 0$ , we have*

$$\tilde{\alpha} = \frac{\sum_j \int \int E \left[ \tau_i^{\delta_j} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j)}{\sum_j \int \int E \left[ \tau_i^{D_j} (x_+, z_+, x_-, z_-) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j)}. \quad (6)$$

The numerator in (6) measures the  $h(j)$ -weighted average causal effect of shifting the distribution of  $(X_i, Z_i)$  from  $H_-(\cdot|j)$  to  $H_+(\cdot|j)$  on  $\delta_j$ , while the denominator measures the weighted average effect of the same change on  $D_{i,j}$ .

This structure is familiar from other instrumental variables settings, and suggests that we may, under further conditions, be able to express  $\tilde{\alpha}$  in terms of causal effects of  $D_i$  on  $\delta$ . Building on Angrist et al. (2000), we show that this is indeed the case provided  $Y_i(\cdot)$  and  $\delta(\cdot)$  are smooth.

**Assumption 6.** *Suppose that  $\delta(y, x; \theta)$  is everywhere continuously differentiable in  $(y, x)$ , while  $Y_i(d, x)$  is everywhere continuously differentiable in  $(d, x)$  almost surely.*

Under this assumption we can define the local causal effects of  $D_i$  and  $X_i$  on  $Y_i$  and  $\delta$ .

**Definition 4.** The local causal effect of  $D_i$  on  $Y_i$  at  $(d, x)$  is  $\mathcal{T}_i^{DY}(d, x) \equiv \frac{\partial}{\partial d} Y_i(d, x)$ , and the local causal effect of  $D_i$  on  $\delta(Y_i, X_i; \tilde{\theta})$  at  $(d, x)$  is

$$\mathcal{T}_i^{D\delta}(d, x; \tilde{\theta}) \equiv \frac{\partial}{\partial y} \delta(Y_i(d, x), x; \tilde{\theta}) \mathcal{T}_i^{DY}(d, x).$$

Analogously, the local causal effect of  $X_i$  on  $Y_i$  at  $(d, x)$  is  $\mathcal{T}_i^{XY}(d, x) \equiv \frac{\partial}{\partial x} Y_i(d, x)$ , and the local causal effect of  $X_i$  on  $\delta(Y_i, X_i; \tilde{\theta})$  at  $(d, x)$  is

$$\mathcal{T}_i^{X\delta} = \frac{\partial}{\partial x} \delta(Y_i(d, x), x; \tilde{\theta}) + \frac{\partial}{\partial y} \delta(Y_i(d, x), x; \tilde{\theta}) \mathcal{T}_i^{XY}(d, x).$$

To compactly express  $\tilde{\alpha}$  in terms of local causal effects, for  $t \in [0, 1]$  let

$$D_i(t, x_+, z_+, x_-, z_-) = t \cdot D_i(x_+, z_+) + (1 - t) \cdot D_i(x_-, z_-)$$

denote the  $t$ -weighted average of  $D_i(x_+, z_+)$  and  $D_i(x_-, z_-)$ , and correspondingly let  $X(t, x_+, x_-) = t \cdot x_+ + (1 - t) \cdot x_-$  denote the  $t$ -weighted average of  $x_+$  and  $x_-$ . For any function  $A_i(d, x)$  of  $(d, x)$ , we use the shorthand

$$A_i(t, x_+, z_+, x_-, z_-) \equiv A_i(D_i(t, x_+, z_+, x_-, z_-), X(t, x_+, x_-)).$$

**Definition 5.** For a random function  $A_i(d, x)$ , define the  $\tau^D$ -weighted linear transformation  $L_D(\cdot)$  as

$$L_D(A_i(\cdot)) \equiv \sum_j \int \int \left( \int_0^1 E \left[ e_j' A_i(t, x_+, z_+, x_-, z_-) \omega_i(x_+, z_+, x_-, z_-) \right] dt \right) dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j)$$

for  $e_j$  the  $j$ th standard basis vector (the vector with a one in the  $j$ th position and zeros elsewhere) and

$$\omega_i(x_+, z_+, x_-, z_-) = \frac{\tau_i^D(x_+, z_+, x_-, z_-)}{\sum_j \int \int E \left[ \tau_i^{Dj}(x_+, z_+, x_-, z_-) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j)}.$$

Correspondingly define the  $\tau^X$ -weighted linear transformation  $L_X(\cdot)$  as

$$L_X(A_i(\cdot)) \equiv \sum_j \int \int \left( \int_0^1 E \left[ e_j' A_i(t, x_+, z_+, x_-, z_-) \zeta_i(x_+, x_-) \right] dt \right) dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j)$$

for

$$\zeta_i(x_+, x_-) = \frac{\tau_i^X(x_+, x_-)}{\sum_j \int \int E \left[ \tau_i^{Dj}(x_+, z_+, x_-, z_-) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j)}.$$

The linear transformations  $L_D(\cdot)$  and  $L_X(\cdot)$  play an important role in our remaining results. Both are linear operators which map random, matrix-valued functions to scalars, and  $L_D$  has the property that  $L_D(I) = 1$ . More generally, for  $a \in \mathbb{R}^J$  and  $Diag(a)$  the matrix with  $a$  on the diagonal and zeros elsewhere, we have that  $L_D(Diag(a)) = \sum_j w_j a_j$  where  $\sum_j w_j = 1$ . It may be tempting to interpret  $L_D(\cdot)$  as a weighted average, but note that  $w_j$ , and more generally  $\omega_i(x_+, z_+, x_-, z_-)$ , may be negative. Moreover, when  $J > 1$ ,  $\omega_i(x_+, z_+, x_-, z_-) \in \mathbb{R}^J$  is vector-valued. We thus cannot in general interpret  $L_D(\cdot)$  as a proper weighted average. Note that, unlike  $L_D(\cdot)$ ,  $L_X(I) \neq 1$  in general so  $L_X(\cdot)$  lacks even an improper weighted average interpretation.

**Proposition 2.** *Under Assumptions 1, 3, 5, 6, if the researcher uses a linear residual and  $\sum_j E[D_{i,j} f_j^v(X_i, Z_i)] \neq 0$ ,*

$$\tilde{\alpha} = L_D \left( \mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta}) \right) + L_X \left( \mathcal{T}_i^{X\delta} (\cdot; \tilde{\theta}) \right). \quad (7)$$

Proposition 2 shows that we can express  $\tilde{\alpha}$  as the sum of two terms. The first term in (7),  $L_D \left( \mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta}) \right)$ , is the  $\tau^D$ -weighted linear transformation of the local causal effect of  $D_i$  on  $\delta$ ,  $\mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta})$ . Interpretability of this term will depend on the structure of  $\delta$  in a given setting. One of our main tasks below is to simplify  $\mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta})$  using the structure of particular estimators.<sup>9</sup> The second term in (7),  $L_X \left( \mathcal{T}_i^{X\delta} (\cdot; \tilde{\theta}) \right)$ , is the  $\tau^X$ -weighted linear transformation of the local causal effect of  $X_i$  on  $\delta$ . A researcher interested in the *ceteris paribus* effect of  $D_i$  might prefer to eliminate this term.

Similar to Corollary 1 above, if the researcher conducts estimation using excluded instruments, the interpretation simplifies and  $\tilde{\alpha}$  reflects only the causal effect of  $D_i$ .

**Corollary 3.** *Under Assumptions 1, 3, 5, 6, if the researcher uses a linear residual and excluded instruments, and if  $\sum_j E[D_{i,j} \tilde{f}_j(Z_i)] \neq 0$ , then*

$$\tilde{\alpha} = L_D \left( \mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta}) \right).$$

Intuitively, when the researcher uses excluded instruments,  $H_+ (\cdot|j)$  and  $H_- (\cdot|j)$  imply the same distribution for  $X_i$ . This implies, however, that  $L_X (A_i (\cdot)) = 0$  for all  $A_i (\cdot)$ , so we eliminate the potentially undesirable term in  $\tilde{\alpha}$ .

---

<sup>9</sup>We can equivalently write  $L_D \left( \mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta}) \right)$  in terms of local causal effects  $\mathcal{T}_i^{DY} (d, x)$  of  $D_i$  on  $Y_i$ ,

$$L_D \left( \mathcal{T}_i^{D\delta} (\cdot; \tilde{\theta}) \right) = L_D \left( \Delta_i (\cdot; \tilde{\theta}) \mathcal{T}_i^{DY} (\cdot) \right)$$

where  $\Delta_i (d, x) = \frac{\partial}{\partial y} \delta (Y_i (d, x), x; \tilde{\theta})$ . While this expression may appear more interpretable,  $\Delta_i (d, x)$  depends on the potential outcomes  $Y_i (\cdot)$ . This renders  $\tilde{\alpha}$  nonlinear in  $\mathcal{T}_i^{DY} (\cdot)$ , since if the potential outcomes change then the “weights”  $\Delta_i (d, x)$  change as well.

## 4 Linear Instrumental Variables Model

As an illustration, we start with an application to linear instrumental variables (IV) estimation. The constant-effect linear IV model implies potential outcomes

$$Y_i^*(d, x, \xi_i; \theta) = \alpha d + x' \beta + \xi_i, \quad (8)$$

while the common IV estimators can be written as GMM with moments  $E[(Y_i - \alpha D_i - X_i' \beta)(X_i', Z_i')'] = 0$ .<sup>10</sup>

In our framework, this corresponds to the case with  $J = \dim(D_i) = 1$  and linear residual

$$R^*(Y_i, D_i, X_i; \theta) = \delta(Y_i, X_i; \theta) - \alpha D_i \text{ for } \delta(Y_i, X_i; \theta) = Y_i - X_i' \beta. \quad (9)$$

In just-identified settings ( $\dim(Z_i) = 1$ ), the IV estimate satisfies Assumption 3 for  $f(X_i, Z_i) = (X_i', Z_i)'$ .<sup>11</sup>

If the researcher estimates the linear IV model using included instruments, Proposition 2 implies that

$$\tilde{\alpha} = L_D (\mathcal{T}_i^{DY}(\cdot)) + L_X (\mathcal{T}_i^{XY}(\cdot) - \tilde{\beta}).$$

Thus,  $\tilde{\alpha}$  reflects both (i) the local causal effect of  $D_i$  on  $Y_i$ ,  $\mathcal{T}_i^{DY}(\cdot)$ , and (ii) the difference between the true and model-implied local causal effect of  $X_i$  on  $Y_i$ ,

$$\mathcal{T}_i^{XY}(\cdot) - \tilde{\beta} = \mathcal{T}_i^{XY}(\cdot) - \frac{\partial}{\partial x} Y_i^*(d, x, \xi_i; \tilde{\theta}).$$

By contrast, if the researcher estimates the linear IV model using excluded instruments, Corollary 3 implies that  $\tilde{\alpha} = L_D (\mathcal{T}_i^{DY}(\cdot))$ .<sup>12</sup>

<sup>10</sup>This is a special case of a constant-effect linear potential outcome model  $Y_i^*(d, x, \xi_i; \theta) = \alpha d + c(x)' \beta + \xi_i$ , where  $c(\cdot)$  is a known, vector-valued function. We focus on the special case with  $c(x) = x$  for the sake of exposition, but our results extend directly to the more general case.

<sup>11</sup>In over-identified settings the two-stage least squares estimate satisfies Assumption 3 for

$$f(X_i, Z_i) = E[(D_i, X_i')' W_i] E[W_i W_i']^{-1} W_i \text{ for } W_i = (X_i', Z_i)'$$

<sup>12</sup>Blandhol et al. (2022) and Słoczyński (2022) analyze the causal interpretation of two stage least squares specifications that use excluded instruments and parametrically control for covariates. Both papers focus on the linear IV model, make the weaker independence assumption that  $X_i, (Y_i(\cdot), D_i(\cdot)) \perp\!\!\!\perp Z_i | X_i$ , and assume that the functional form for  $X_i$  is in a sense correctly specified. Under these assumptions, these papers analyze the impact of different ways of controlling for  $X_i$  on the interpretation of  $\tilde{\alpha}$  in terms of causal effects of  $D_i$ . By contrast we impose the stronger independence condition  $(Y_i(\cdot), D_i(\cdot)) \perp\!\!\!\perp Z_i \perp\!\!\!\perp X_i$ , make no assumption on the functional form for  $X_i$ ,

These results establish a sense in which the researcher’s estimate is sensitive to misspecification of the causal relationship between  $X_i$  and  $Y_i$  when using included instruments but not when using excluded instruments. To illustrate this contrast, for a given distribution of potential outcomes  $Y_i(\cdot)$ , define new potential outcomes  $Y_i^q(d, x) = Y_i(d, x) + q(x)$  which add an arbitrary function  $q(x)$ . By definition  $L_D(\mathcal{T}_i^{DY}(\cdot)) = L_D(\mathcal{T}_i^{DY^q}(\cdot))$ , so  $\tilde{\alpha}$  is unaffected by this change under excluded instruments. By contrast,  $L_X(\mathcal{T}_i^{XY}(\cdot)) \neq L_X(\mathcal{T}_i^{XY^q}(\cdot))$  in general, and  $\tilde{\beta}$  will also change, so under included instruments  $\tilde{\alpha}$  will be sensitive to the change from  $Y_i(\cdot)$  to  $Y_i^q(\cdot)$ . Assumption 1(b) is important for this conclusion, since if  $Z_i$  were not independent of  $X_i$ , the estimand under excluded instruments would be affected by the addition of  $q(x)$  to the potential outcomes.

**Special Case: Interaction Instruments** A special case of an included instrument is one that consists of the product of two (or more) elements of  $X_i$ . For example, Gentzkow and Shapiro (2008) estimate the effect of child  $i$ ’s exposure to television  $D_i$  on their score  $Y_i$  from a standardized test taken in adolescence. Gentzkow and Shapiro (2008, equations 1 and 2) use as an instrument the product of indicators for grade (cohort) and indicators for when the child’s home market adopted television, while including both grade and market indicators directly in their outcome model  $Y^*(\cdot)$ .<sup>13</sup> ■

#### 4.1 Numerical Illustration

Suppose that  $Y_i$ ,  $D_i$ ,  $X_i$ , and  $Z_i$  are scalars and that the true data-generating process satisfies Assumption 1 with

$$Y_i(d, x) = d + x + \gamma(\ln(x) - x) + \xi_i,$$

so that the local causal effect of  $D_i$  on  $Y_i$  is equal to one, i.e.,  $\mathcal{T}_i^{DY}(d, x) = 1$  for all  $(d, x)$  and all  $i$ . We consider  $\gamma \in [0, 1]$ .

---

and focus on the distinction between excluded and included instruments.

<sup>13</sup>A different sort of interaction instrument interacts an excluded instrument with functions of  $X_i$ ,  $f(X_i, Z_i) = Z_i f(X_i)$ . Note that in our framework,  $Z_i^* = Z_i f(X_i)$  is generally an included instrument, since while  $(Y_i(\cdot), D_i(\cdot)) \perp\!\!\!\perp Z_i^*$ , in general  $Z_i^* \not\perp\!\!\!\perp X_i$ .



We mimic a researcher who does not know the true model and instead specifies a model satisfying Assumption 2 with

$$Y^*(d, x, \xi; \theta) = \mu + \alpha d + x\beta + \xi_i.$$

Thus when  $\gamma = 0$  the researcher's model is consistent with the true model, but when  $\gamma > 0$  it is not. The value of  $\gamma$  can be taken as a measure of the degree of misspecification.

We suppose that the researcher estimates the parameter  $\alpha$ , which describes the causal effect of  $D_i$  on  $Y_i$  under the researcher's model, using a linear IV estimator with residual (9). We consider three sets of instruments: one (excluded instruments) that includes  $Z_i$ , one (included instruments) that replaces  $Z_i$  with  $X_i^2$ , and one (endogenous variable as instrument) that replaces  $Z_i$  with  $D_i$ .

Figure 2 plots the estimated median bias of the researcher's estimate of the effect of  $D_i$  on  $Y_i$  as a function of the degree of misspecification  $\gamma$  for each choice of instrument. When  $\gamma = 0$  and so the researcher's model is correct, the estimator is approximately median-unbiased when based on either excluded or included instruments, but severely median-biased when based on using the endogenous variable as the instrument. When  $\gamma > 0$  and so the researcher's model is incorrect, the estimator is approximately median-unbiased when based on excluded instruments but can be severely median-biased when based on included instruments or using the endogenous variable as the instrument. When  $\gamma$  is sufficiently large, the estimator based on included instruments has greater median bias than the estimator using the endogenous variable as the instrument.

## 5 Differentiated Goods Demand Model with Price Endogeneity

We next turn to an application where the researcher adopts a nonlinear model, specifically a random coefficients logit demand model. In this setting the researcher observes data from  $n$  markets indexed by  $i$ . In each market there are  $J + 1$  products, with  $j = 0$  an outside option.  $Y_i$  is the vector of market shares,  $D_i$  a vector of prices,  $X_i$  a matrix of other characteristics (assumed to include a constant), and  $Z_i$  a matrix of cost shifters.

The researcher assumes that a unit mass of consumers  $c$  in the market each choose one product  $j$  to maximize utility  $u_{c,i,j}$  given by

$$u_{c,i,j} = \phi_{i,j} + X'_{i,j} \Upsilon \psi_{c,i} + \epsilon_{c,i,j}, \text{ for } \phi_{i,j} = X'_{i,j} \beta + \alpha D_{i,j} + \xi_{i,j} \quad (10)$$

where  $\psi_{c,i}$  is an i.i.d. mean-zero random coefficient with a known distribution and  $\epsilon_{c,i,j}$  is a

consumer-specific utility shock that follows an i.i.d. type 1 extreme value distribution independently of all other variables, so that  $\phi_{i,j}$  describes the mean net-of-price utility of good  $j$  in market  $i$ . The model in (10) is similar to the one considered in Berry et al. (1995) but does not allow a random coefficient on the price  $D_{i,j}$ . Berry (1994) shows that for known  $\Upsilon$  we can recover  $\phi_{i,j}$  as a function of observable market shares and characteristics,  $\phi_{i,j} = \phi_j(Y_i, X_i; \Upsilon)$ , and Berry et al. (1995) use this observation to construct GMM estimators based on a residual  $R_j^*(Y_i, D_i, X_i; \theta)$ , which under (10) takes the form

$$R_j^*(Y_i, D_i, X_i; \theta) = \delta_j(Y_i, X_i; \theta) - \alpha D_{i,j}, \text{ for } \delta_j(Y_i, X_i; \theta) = \phi_j(Y_i, X_i; \Upsilon) - X'_{i,j}\beta. \quad (11)$$

Other models with residuals of the form in (11) are discussed in, for example, Berry (1994).

If the researcher estimates (10) using excluded instruments, Corollary 3 implies that the limiting price coefficient  $\tilde{\alpha}$  may be interpreted in terms of the local causal effect  $\mathcal{T}_i^{D\delta}$  of prices  $D_i$  on  $\delta_j(Y_i, X_i; \tilde{\theta})$ . Moreover, since  $D_i$  has no causal effect on  $X_i$ , the causal effect of  $D_i$  on  $\delta(Y_i, X_i; \tilde{\theta})$  is the same as the causal effect on the model-implied mean utility,  $\tilde{\alpha} = L_D(\mathcal{T}_i^{D\phi}(\cdot; \tilde{\theta}))$  for  $\mathcal{T}_i^{D\phi}(d, x; \tilde{\theta})$  the local causal effect of  $d$  on  $\phi(Y_i, X_i; \tilde{\Upsilon})$  at  $(d, x)$ .

Unfortunately, the interpretation of  $\tilde{\alpha}$  in terms of local causal effects  $\mathcal{T}_i^{D\phi}$  is not especially illuminating. The model-implied mean utility  $\phi(Y_i, X_i; \Upsilon)$  is a complex nonlinear transformation of the data, and is in general difficult to interpret under model misspecification. Fortunately, the structure of the model (10) allows us to go further, and directly relate the true and model-implied causal effects of price in this setting. To state this result, let  $\mathcal{T}_i^{DY^*}(d, x; \tilde{\theta})$  denote the local causal effect of  $d$  on  $Y_i$  at  $(d, x)$  implied by the model (10).

**Proposition 3.** *Under Assumptions 1, 3, 5, and 6, if the researcher uses residual function (10) with excluded instruments,  $\sum_j E[D_{i,j}\tilde{f}_j(Z_i)] \neq 0$ , and  $\mathcal{T}_i^{DY^*}(d, x; \tilde{\theta})$  is invertible almost surely, then*

$$L_D\left(\mathcal{T}_i^{DY^*}(d, x; \tilde{\theta})^{-1}\mathcal{T}_i^{DY}(d, x)\right) = 1. \quad (12)$$

This result shows that there is a nonstandard sense in which the model-implied causal effects match the true causal effects on average, where the true and model-implied causal effects enter in a nonseparable way. Under correct specification and identification of the researcher's model (i.e., Assumptions 2 and 4),  $\mathcal{T}_i^{DY^*}(d, x; \tilde{\theta}) = \mathcal{T}_i^{DY}(d, x)$  for all  $i$ , so (12) holds trivially. Proposition 3 establishes that (12) continues to hold even under misspecification, provided the researcher uses

excluded instruments. Note that Proposition 3 holds for any demand system such that (i) the residual takes the form (11) and (ii)  $Y_i = Y^*(d, x, \xi; \theta_0)$  is a one-to-one differentiable function of  $\xi$  for all  $d, x$ , and  $\theta$ , whether or not utility takes the form (10).

**Special Case: Multinomial Logit** To build further intuition, consider the special case in which the researcher assumes that there is no random coefficient,  $\Upsilon = 0$ , so (10) reduces to the multinomial logit. Proposition 3 continues to hold in this setting, but the model-implied mean utility simplifies to  $\phi_j(Y_i, X_i; 0) = \log(Y_{i,j}) - \log(Y_{i,0})$ , so  $\tilde{\alpha}$  has a more direct economic interpretation.

**Corollary 4.** *Under Assumptions 1, 3, 5, and 6, if the researcher uses residual function (10) with  $\Upsilon = 0$  and excluded instruments, and  $\sum_j E[D_{i,j} \tilde{f}_j(Z_i)] \neq 0$ , then  $\tilde{\alpha} = L_D(\Delta S_i(\cdot))$  where  $\Delta S_{i,j}(d, x)$  is the semi-elasticity of share  $j$  with respect to the vector of prices, less the semi-elasticity of the outside option,*

$$\Delta S_{i,j}(d, x) = \frac{\frac{\partial}{\partial d} Y_{i,j}(d, x)}{Y_{i,j}(d, x)} - \frac{\frac{\partial}{\partial d} Y_{i,0}(d, x)}{Y_{i,0}(d, x)}.$$

Corollary 4 establishes that, in the multinomial logit estimated with excluded instruments,  $\tilde{\alpha}$  is determined by the semi-elasticity of the inside goods with respect to price, relative to the semi-elasticity of the outside good. Note that when there are multiple inside goods,  $J > 1$ ,  $\Delta S_{i,j}(d, x)$  is a  $1 \times J$  vector measuring semi-elasticities with respect to all prices, not just the own-price semi-elasticity. Hence, in general  $\tilde{\alpha}$  reflects a composite of many different types of causal effects.

■

A recent literature studies conditions under which differentiated goods demand models admit a representation in which a product-specific mean utility can be inverted from market shares (see, e.g., Berry et al. 2013; Berry and Haile 2014). In general such conditions are not sufficient to guarantee that  $\tilde{\alpha}$  depends only on causal effects of prices when included instruments are used. The reason is that the existence, under the true model, of some residual function of the form in (11) is not sufficient to guarantee that this residual function takes the specific form assumed by the researcher. Intuitively, nonparametric identification results are not enough to ensure that parametric estimators based on (incorrect) functional form assumptions capture causal effects of price.<sup>14</sup>

<sup>14</sup>The Berry and Haile (2014) model implies that, for  $X_{i,j} = (X_{i,j}^{(1)}, X_{i,j}^{(2)})$ , potential outcomes take the form

$$Y_i(d, x) = \tilde{\sigma}(x^{(1)} + \xi_i, x^{(2)}, d)$$

## 5.1 Application to the Demand for Beer

Our formal results establish a sense in which the researcher’s estimand can be interpreted in terms of the *ceteris paribus* effect of  $D_i$  on  $Y_i$  when the researcher uses excluded instruments but not when the researcher uses included instruments. Even under excluded instruments, however, the researcher’s estimand need not correspond to an object of economic interest, so our theoretical results need not imply that using excluded instruments improves recovery of the objects of primary interest to the researcher. In this subsection we investigate how instrument choice affects the recovery of objects of economic interest in a simulation exercise calibrated tightly to an economic application.

Miller and Weinberg (2017) estimate a differentiated-goods model of the demand for beer in the United States using data from the IRI Academic Database (Bronnenberg et al. 2008). In Miller and Weinberg’s setting, a market  $i$  is a region-month, the outcome  $Y_i$  is the vector of market shares of  $J$  products, the endogenous variable  $D_i$  is the vector of prices of the  $J$  products, the matrix  $X_i$  describes the characteristics and ownership of each of the  $J$  products, and the excluded variable  $Z_i$  is a vector of cost shifters for the  $J$  products, constructed by multiplying an index of diesel prices by the distance of the market from the product’s brewery.

Miller and Weinberg (2017) specify a random-coefficients nested logit demand model with a  $J$ -dimensional structural residual  $\xi_i$ . Miller and Weinberg (2017) also specify a Bertrand-Nash model of beer pricing with a  $J$ -dimensional structural residual  $\eta_i$ . Miller and Weinberg (2017) estimate their demand model using GMM, adopting the nested fixed point procedure of Berry et al. (1995). Miller and Weinberg (2017) likewise estimate their pricing model using GMM, with moments based on firms’ first-order conditions.<sup>15</sup>

To explore the effects of different forms of misspecification under different choices of instruments, we treat Miller and Weinberg’s (2017) estimated demand model as the true potential outcome function  $Y_i(\cdot)$  and Miller and Weinberg’s (2017) estimated pricing model as the true potential endogenous variable function  $D_i(\cdot)$ . We simulate 100 replicates of data generated from the model by independently permuting  $\xi_{ij}$ ,  $\eta_{ij}$ , and  $Z_{ij}$  across all  $i, j$  pairs and then recomputing the implied

---

where  $\tilde{\sigma}$  is invertible in its first argument and  $\xi_i$  is a  $J$ -dimensional random vector. Under these conditions one can re-express our results in terms of the objects Berry and Haile (2014) study, noting that, e.g.,  $\frac{\partial}{\partial d} Y_i(d, x) = \frac{\partial}{\partial d} \tilde{\sigma}(x^{(1)} + \xi_i, x^{(2)}, d)$ . This structure, in conjunction with Berry and Haile’s (2014) other assumptions, suffices to identify  $\xi_i$ . If, however, the researcher’s parametric assumptions in (11) are incorrect, then the model-implied residuals  $\delta(Y_i(D_i(x, z), x; \tilde{\theta}))$  will generally differ from  $\xi_i$ , and  $\hat{\alpha}$  estimated using included instruments will generally capture both causal effects of prices and the direct effect of the included instruments.

<sup>15</sup>See Miller and Weinberg (2017), equations 2, 9, 8, and 13, respectively.

market shares  $Y_i$  and prices  $D_i$  in each market  $i$ .

We mimic a researcher who does not know the true model and instead specifies some, possibly wrong, model  $Y^*(\cdot)$  that features a linear residual and that can be estimated via GMM. For each model  $Y^*(\cdot)$  that we specify, we estimate the model via exactly identified GMM using three sets of instruments, one (excluded instruments) that contains  $Z_{ij}$ , one (included instruments) that replaces  $Z_{ij}$  with the number  $N_{ij} = N_j(X_i)$  of products available in market  $i$  that are owned by firms other than the owner of  $j$ , and one (price as instrument) that replaces  $Z_{ij}$  with  $D_{ij}$ . Each model  $Y^*(\cdot)$  that we consider features at least one product-specific parameter and at least one common parameter.

We study recovery of the average own-price elasticity, a possible measure of market power. Our theoretical results do not imply that the estimators we consider will recover the true average own-price elasticity when the researcher's model  $Y^*(\cdot)$  does not coincide with the true model  $Y(\cdot)$ . However, the contrast we find in Section 3.1 between the interpretation of the estimand under excluded and included instruments leads us to hope that the estimator will tend to be less biased when based on excluded instruments.

Figure 3 presents our findings. For each model  $Y^*(\cdot)$  and each choice of instruments, we present the estimated median bias of the average own-price elasticity based on the 100 replicates. To describe sampling uncertainty in the median bias due to our use of a finite number of simulation replicates, we include a 95 percent confidence interval for the median bias, but this confidence interval is typically narrower than the markers we use to display results, and so is typically not visible.

When the researcher's model  $Y^*(\cdot)$  corresponds to the true model  $Y(\cdot)$ , we find that the estimator is approximately median-unbiased when based on either excluded or included instruments and severely median-biased when using price as instrument. The mean elasticity under Miller and Weinberg's (2017) model and data is  $-4.77$ . When using excluded instruments and the correct specification, the estimated median bias is  $-0.03$ . When using included instruments and the correct specification, the estimated median bias is  $-0.08$ . When using price as instrument and the correct specification, the estimated median bias is  $1.50$ .

When the researcher's model  $Y^*(\cdot)$  does not correspond to the true model  $Y(\cdot)$ , we find that the estimator remains approximately median-unbiased when based on excluded instruments but can be severely median-biased when based on included instruments or price as instrument. When using excluded instruments, the estimated median bias ranges from  $-0.10$  to  $-0.03$  across all

specifications other than the correct specification. When using included instruments, the estimated median bias ranges from  $-5.23$  to  $2.46$ . When using price as instrument, the estimated median bias ranges from  $1.32$  to  $2.31$ . In this setting and for these forms of misspecification, then, estimators based on included instruments can have greater median bias than estimators based on ignoring price endogeneity, whereas estimators based on excluded instruments are approximately median-unbiased.

Appendix C contains additional simulation results. Appendix Figure 1 reports estimates of the coverage of 95 percent confidence intervals for the average own-price elasticity. Appendix Figure 2 reports estimates of the median bias for the median own-price elasticity rather than the average own-price elasticity. Appendix Figure 3 reports estimates of the median bias for the average own-price elasticity, including estimates we exclude from our main analysis due to estimated parameters hitting a boundary. Appendix Figure 4 reports estimates of the median bias including additional specifications  $Y^*(\cdot)$  that do not exhibit a linear residual and are therefore covered by our general results in Section 3 but not those in Section 3.1.

## 6 Dynamic Extension: Production Model with Input Endogeneity

In this section we extend our analysis to cover dynamic settings, focusing on dynamic panel approaches to production function estimation. Here  $i$  indexes firms,  $j$  indexes time periods,  $Y_i$  is a vector of log outputs,  $D_i$  is a vector of log quantities for a static input, and  $Z_i$  collects a sequence of input cost shifters. The covariates  $X_{i,j}$  consist of state variables including past values  $Y_{i,1:j-1} = \{Y_{i,1}, \dots, Y_{i,j-1}\}$  of output, past values  $D_{i,1:j-1} = \{D_{i,1}, \dots, D_{i,j-1}\}$  of the static input, and past and current values  $K_{i,1:j} = \{K_{i,1}, \dots, K_{i,j}\}$  of a dynamic input.

The researcher assumes that production is governed by a Cobb-Douglas technology with

$$\begin{aligned} Y_{i,j} &= \mu + \alpha D_{i,j} + \beta K_{i,j} + \zeta_{i,j} \\ \zeta_{i,j} &= \nu \zeta_{i,j-1} + \nu_{i,j} \text{ for } j > 0 \end{aligned} \tag{13}$$

where  $\mu$  is a constant and  $\zeta_{i,0}$  is drawn from some distribution. Here  $\zeta_{i,j}$  is productivity, and evolves as an AR(1) process with innovation  $\nu_{i,j}$ , where  $\nu_{i,j}$  is independent over time with  $E[\nu_{i,j}] = 0$ . The innovation  $\nu_{i,j}$  is realized after the dynamic input is chosen but before the static input is chosen in period  $j$ , and is therefore independent of  $X_{i,j}$ , but not of  $D_{i,j}$  or  $X_{i,j+1}$ . As discussed in Akerberg

et al. (2015, section 4.3.3), under standard assumptions this model implies that we may conduct GMM estimation based on the residual function

$$R_j^*(Y_i, D_i, X_i; \theta) = (Y_{i,j} - vY_{i,j-1}) - \mu(1 - v) - \alpha(D_{i,j} - vD_{i,j-1}) - \beta(K_{i,j} - vK_{i,j-1}). \quad (14)$$

To analyze performance under misspecification, we again nest the researcher's model in a potential outcomes framework. To accommodate the dynamic structure of this setting, we modify Assumption 1 to reflect the timing of decisions in the model.

**Assumption 7.** (*Static input*) Under the true data-generating process the following hold:

- (a) (*dynamic exclusion*) Assumption 1(a) holds and, further,  $D_{i,j}(x, z) = D_{i,j}(x, z')$  for all  $x \in \mathcal{X}$  and all  $z, z' \in \mathcal{Z}$  such that  $z_j = z'_j$ .
- (b) (*random assignment*)  $(Y_{i,j}(\cdot), D_{i,j}(\cdot), X_{i,j}) \perp\!\!\!\perp Z_{i,j}$  and  $Z_{i,j} \perp\!\!\!\perp Z_{i,k}$  for  $j \neq k$ .

Assumption 7(a) states that only the contemporaneous cost shifters impact the choice of the static input in a given period, while Assumption 7(b) states that the unobservable factors affecting output and the static input in a given period are independent of the state variables and instruments in that period, and that the instrument is independent across time. Importantly, Assumption 7 allows shocks to output, e.g. productivity shocks, to influence the level of the static input, and thus for econometric endogeneity in a naive regression of output on inputs (Marschak and Andrews 1944). It also allows for an arbitrary relationship between the state variables in a given period and past values of the instrument, consistent with the fact that the state variables include past values of output and of dynamically chosen inputs.

Again letting  $\tilde{\theta}$  denote the researcher's estimand, we are interested in the interpretation of the coefficient  $\tilde{\alpha}$  on the static input. Note that (14) is a linear residual.<sup>16</sup> Provided the researcher uses *excluded, contemporaneous instruments*,  $f_j(X_i, Z_i) = \tilde{f}_j(Z_{i,j})$ , we obtain an expression for  $\tilde{\alpha}$  in terms of the local causal effect of the static input,  $\mathcal{T}_{i,j}^{D_j Y}(\cdot) = \frac{\partial}{\partial d_j} Y_{i,j}(d, x)$ .

**Proposition 4.** Under Assumptions 3, 5, 6, and 7, if the researcher uses *excluded, contemporaneous instruments*, and if  $\sum_j E[D_{i,j} \tilde{f}_j(Z_i)] \neq 0$ , then we have that  $\tilde{\alpha} = L_D^* \left( \mathcal{T}_{i,j}^{D_j Y}(\cdot) \right)$ , where for

<sup>16</sup>Specifically,

$$\delta(Y_i, X_i; \theta) = Y_{i,j} - vY_{i,j-1} - \mu(1 - v) + \alpha v D_{i,j-1} - \beta(K_{i,j} - vK_{i,j-1}).$$

any scalar-valued  $A_{i,j}(d, x)$ ,

$$L_D^*(A_{i,j}(\cdot)) =$$

$$\sum_j \int_{\mathcal{Z}} \int_{\mathcal{Z}} \left( \int_0^1 E[A_{i,j}(D_i(t, X_i, z_+, z_-), X_i) \omega_{i,j}(X_i, z_+, z_-)] dt \right) dH_+(z_+|j) dH_-(z_-|j) h(j)$$

for  $D_i(t, X_i, z_+, z_-) = t \cdot D_i(X_i, z_+) + (1-t) \cdot D_i(X_i, z_-)$  and

$$\omega_{i,j}(X_i, z_+, z_-) = \frac{\tau_i^{D_j}(X_i, z_+, z_-)}{\sum_j \int \int E[\tau_i^{D_j}(X_i, z_+, z_-)] dH_+(z_+|j) dH_-(z_-|j) h(j)}.$$

Note that both  $Y_{i,j}$  and  $D_{i,j}$  are measured in logs, so  $\mathcal{T}_{i,j}^{D_j Y}(\cdot)$  can be interpreted as an elasticity. Hence, when the researcher uses excluded, contemporaneous instruments, the limiting coefficient  $\tilde{\alpha}$  on the static input recovers a linear transformation of the output elasticity with respect to the contemporaneous static input. The weights  $\omega_{i,j}(X_i, z_+, z_-)$  average to one and are scalar, unlike in the general case.<sup>17</sup>

If the researcher instead uses included instruments, then analogous to Proposition 2 the estimand  $\tilde{\alpha}$  reflects causal effects of both  $d$  and  $x$ . Interestingly, if the researcher uses non-contemporaneous excluded instruments, e.g.,  $\tilde{f}_j(Z_i)$  which depends on  $Z_{i,j-1}$ ,  $\tilde{\alpha}$  will typically capture causal effects of both  $d$  and  $x$ , since  $Z_{i,j-1}$  can affect  $X_j$ , for example through  $D_{i,j-1}$ .

Gandhi et al. (2020) study nonparametric identification of production functions, maintaining the assumption of Hicks-neutrality, i.e., that  $Y_{i,j}(D_{i,j}, X_{i,j}) = g(D_{i,j}, K_{i,j}; \theta) + \zeta_{i,j}$  for some function  $g(\cdot)$ . Imposing this assumption simplifies the form of causal effects that we study, but does not ensure interpretability of the estimand based on included instrument in terms of causal effects of the input, as it again does not ensure that the researcher's specification of the residual (14) is correct.

## 7 Conclusion

We show theoretically that, under misspecification, a researcher's estimand maintains an interpretation in terms of causal effects of the endogenous variable under excluded instruments but not under included instruments. We show numerically that, under misspecification, a researcher using excluded instruments obtains a less biased estimate of an object of economic interest in a

---

<sup>17</sup>If we further assume that  $D_{i,j}(x, z) > D_{i,j}(x, z')$  for all  $x \in \mathcal{X}$  and all  $z, z' \in \mathcal{Z}$  such that  $z_j < z'_j$ , and the researcher chooses  $f_j(\cdot)$  to be weakly monotone in  $z_j$ , then the weights  $\omega_{i,j}(X_i, z_+, z_-)$  are nonnegative.



simulation calibrated tightly to an economic application.

We recommend that researchers prioritize excluded instruments over included instruments. When excluded instruments are not available, we think our results make it even more important than usual for the researcher to justify their model specification on economic grounds. When such a justification is not available, we think our results make it even more important than usual for the researcher to exhibit the sensitivity of their economic conclusions to other *a priori* reasonable model specifications.

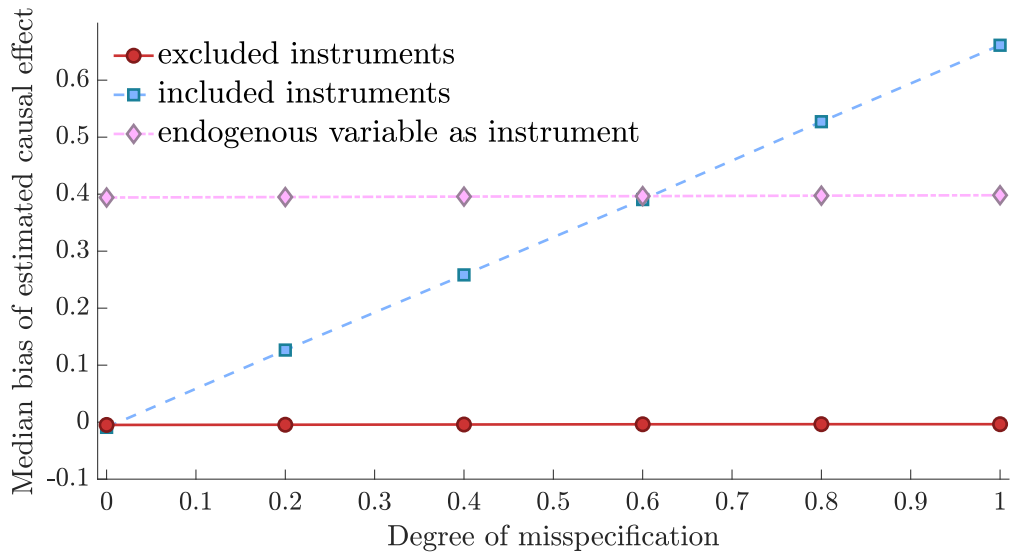
## References

- Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. Econometric tools for analyzing market outcomes. In J. Heckman, E. Leamer (eds.), *Handbook of Econometrics, Volume 6 Part A*: 4171-4276. Amsterdam: North Holland.
- Ackerberg, Daniel, Gregory S. Crawford, and Jinyong Hahn. 2011. Orthogonal instruments: Estimating price elasticities in the presence of endogenous product characteristics. Presentation at CREST, Paris. Slides accessed online at <https://pdfs.semanticscholar.org/b4a6/ff21a06b7364564d929a8a58100e074916bc.pdf> in January 2022.
- Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. Identification properties of recent production function estimators. *Econometrica* 83(6): 2411-2451.
- Andrews, Isaiah. 2019. On the structure of IV estimands. *Journal of Econometrics* 211(1): 294-307.
- Angrist, Joshua D. and Alan B. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106(4): 979-1014.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434): 444-455.
- Angrist, Joshua D., Kathryn Graddy, and Guido W. Imbens. 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies* 67(3): 499-527.
- Angrist, Joshua D. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business & Economic Statistics* 19(1): 2-16.
- Armstrong, Timothy B. 2016. Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica* 84(5): 1961-1980.
- Armstrong, Timothy B. and Michal Kolesár. 2021. Sensitivity analysis using approximate moment condition models. *Quantitative Economics* 12(1): 77-108.
- Berry, Steven T. 1994. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25(2): 242-262.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63(4): 841-890.
- Berry, Steven, Amit Gandhi, and Philip Haile. 2013. Connected substitutes and invertibility of demand. *Econometrica* 81(5): 2087-2111.
- Berry, Steven T. and Philip A. Haile. 2014. Identification in differentiated products markets using market level data. *Econometrica* 82(5): 1749-1797.

- Berry, Steven and Philip Haile. 2016. Identification in differentiated products markets. *Annual Review of Economics* 8:27-52.
- Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky. 2022. When is TSLS actually LATE? *NBER Working Paper No. 29709*.
- Blundell, Richard and Stephen Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87(1): 115-143.
- Blundell, Richard and Stephen Bond. 2000. GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19(3): 321-340.
- Bonhomme, Stéphane and Martin Weidner. Forthcoming. Minimizing sensitivity to model misspecification. *Quantitative Economics*.
- Bresnahan, Timothy F. 1996. Comment on Hausman, “Valuation of new goods under perfect and imperfect competition”. In T. F. Bresnahan, R. J. Gordon (eds.), *The Economics of New Goods: 237-247*. Chicago: University of Chicago Press.
- Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela. 2008. Database paper: The IRI marketing data set. *Marketing Science* 27(4) 745-748.
- Chalak, Karim. 2017. Instrumental variables methods with heterogeneity and mismeasured instruments. *Econometric Theory* 33(1): 69-104.
- Chamberlain, Gary. 1987. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34(1): 305-334.
- de Roux, Nicolás, Marcela Eslava, Santiago Franco, and Eric Verhoogen. 2021. Estimating production functions in differentiated-product industries with quantity information and external instruments. *NBER Working Paper No. 28323*.
- Gandhi, Amit, Salvador Navarro, and David A. Rivers. 2020. On the identification of gross output product functions. *Journal of Political Economy* 128(8): 2973-3016.
- Gandhi, Amit and Jean-François Houde. 2020. Measuring substitution patterns in differentiated-products industries. *NBER Working Paper No. 26375*.
- Gentzkow, Matthew and Jesse M. Shapiro. 2008. Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study. *Quarterly Journal of Economics* 123(1): 279-323.
- Hall, Alastair R. and Atsushi Inoue. 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114(2): 361-394.
- Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50(4): 1029-1054.
- Imbens, Guido W. and Joshua D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62(2): 467-475.
- Kitamura, Yuichi, Taisuke Otsu, and Kirill Evdokimov. 2013. Robustness, infinitesimal neighbor-

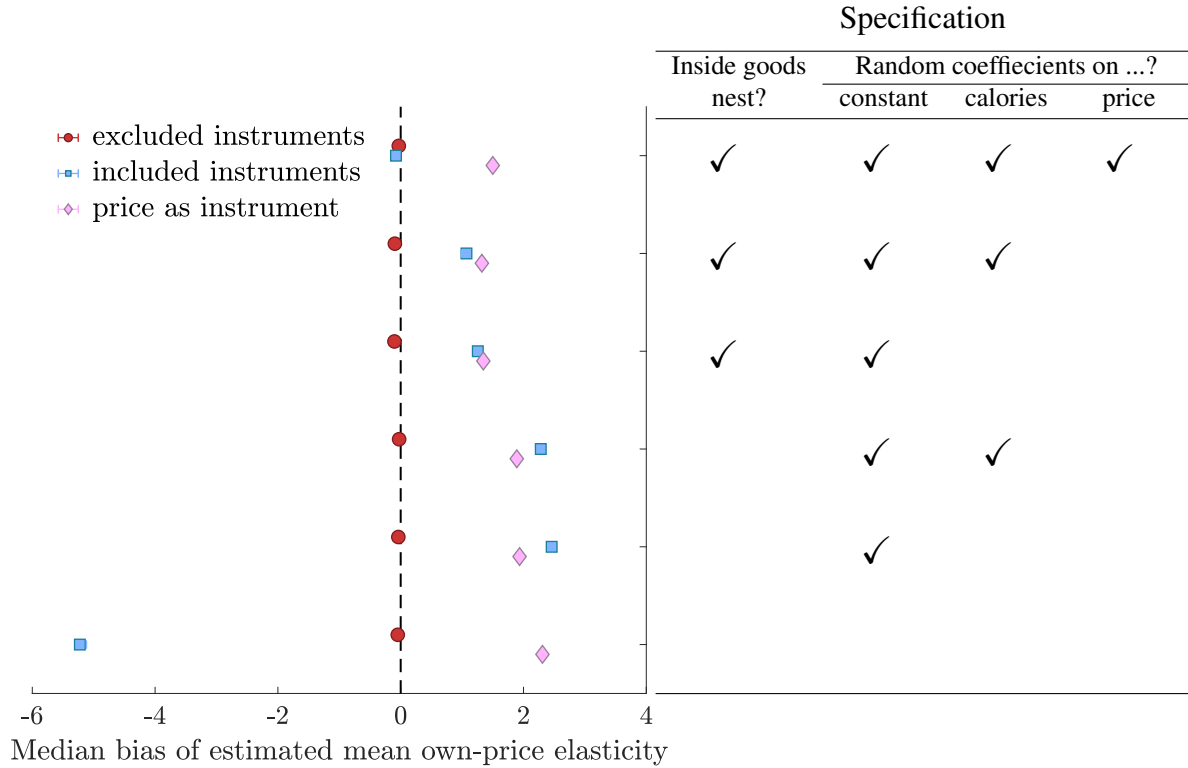
- hoods, and moment restrictions. *Econometrica* 81(3): 1185-1201.
- Kline, Patrick and Christopher R. Walters. 2019. On Heckits, LATE, and numerical equivalence. *Econometrica* 87(2): 677-696.
- Kolesár, Michal. 2013. Estimation in an instrumental variables model with treatment effect heterogeneity. Working paper. Accessed at [https://www.princeton.edu/~mkolesar/papers/late\\_estimation.pdf](https://www.princeton.edu/~mkolesar/papers/late_estimation.pdf) in January 2022.
- Lewbel, Arthur. 2019. The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature* 57(4): 835-903.
- Marschak, Jacob and William H. Andrews. 1944. Random simultaneous equations and the theory of production. *Econometrica* 12(3/4): 143-205.
- Miller, Nathan H. and Matthew C. Weinberg. 2017. Understanding the price effects of the Miller-Coors joint venture. *Econometrica* 85(6): 1763-1791.
- Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. Using instrumental variables for inference about policy relevant treatment variables. *Econometrica* 86(5): 1589-1619.
- Mood, Alexander M. and Franklin A. Graybill. 1963. *Introduction to the Theory of Statistics, Second Edition*. New York: McGraw-Hill.
- Newey, Whitney K. 1990. Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58(4): 809-837.
- Newey, Whitney K. and Daniel McFadden. 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4: 2111-2245.
- Nevo, Aviv. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2): 307-342.
- Nevo, Aviv. 2004. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy* 9(4): 513-548.
- Petrin, Amil and Boyoung Seo. 2019. Identification and estimation of discrete choice demand models when observed and unobserved characteristics are correlated. Working paper. Accessed at <https://economics.smu.edu.sg/sites/economics.smu.edu.sg/files/economics/pdf/Seminar/2019/20190816.pdf> in January 2022.
- Reynaert, Mathias and Frank Verboven. 2014. Improving the performance of random coefficients demand models: The role of optimal instruments. *Journal of Econometrics* 179(1): 83-89.
- Rossi, Peter E. 2014. Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Science* 33(5): 655-672.
- Słoczyński, Tymon. 2022. When Should We (Not) Interpret Linear IV Estimands as LATE? Working paper. Accessed at <https://arxiv.org/abs/2011.06695> in March 2022.

Figure 2: Estimated median bias for estimators of the causal effect of the endogenous variable in a linear IV model



Note: The plot shows the estimated median bias for estimators of the effect of the endogenous variable  $D_i$  on the outcome  $Y_i$  as a function of the degree of misspecification  $\gamma$  in the example in Section 4.1. We estimate the median bias from a simulation with 100 replicates. In each replicate we draw  $(\xi_i, \omega_i) \stackrel{iid}{\sim} N(0, \sigma^2 I)$  with  $\sigma = 0.3$ ,  $Z_i \stackrel{iid}{\sim} U(0, 1)$ , and  $X_i - 1 \stackrel{iid}{\sim} U(0, 1)$ , and calculate the endogenous variable  $D_i = 0.1Z_i + 5\sqrt{X_i} + 0.5\xi_i + \omega_i$  and the outcome  $Y_i = D_i + X_i + \gamma(\ln(X_i) - X_i) + \xi_i$  for each of  $10^5$  units  $i$  and each of the given values of  $\gamma$ . For each replicate and each value of  $\gamma$ , we estimate the effect of  $D_i$  on  $Y_i$  by linear instrumental variables based on the residual  $Y_i - \mu - \alpha D_i - \beta X_i$  implied by the researcher's model. We use three sets of instruments, each of which includes  $(1, X_i)$ . The first set of instruments (excluded instruments) includes  $Z_i$ , the second set of instruments (included instruments) replaces  $Z_i$  with  $X_i^2$ , and the third set of instruments (endogenous variable as instrument) replaces  $Z_i$  with  $D_i$ .

Figure 3: Estimated median bias for estimators of the average own-price elasticity in a differentiated-goods demand model



Note: The plot reports the estimated median bias for estimators of the average own-price elasticity in the setting of Miller and Weinberg (2017) based on 100 simulations. Each row corresponds to a different demand model specification and each marker shape corresponds to a different choice of instruments. We assume that the market shares  $Y_i$  in each market  $i$  are generated from the estimated demand model in Miller and Weinberg (2017, Table IV, column (ii)) and extract the implied shock  $\xi_i$  to preferences in each market. We assume that the prices  $D_i$  are generated from the estimated pricing model in Miller and Weinberg (2017, Table VI, column (ii)) and extract the implied shock  $\eta_i$  to costs in each market. To compute each replicate, we independently permute the elements of  $(\xi_{ij}, \eta_{ij}, Z_{ij})$ , where  $j$  indexes products and  $Z_{ij}$  is a scalar excluded instrument given by the product of a diesel fuel price index with the distance of the market  $i$  from the producer's brewery (Miller and Weinberg 2017, p. 1775). Following each permutation we recompute the market shares  $Y_i$  and prices  $D_i$  implied by the model. For each permutation and each demand model specification, we estimate the specified demand model using three sets of instruments, one (excluded instruments) that contains  $Z_{ij}$ , one (included instruments) that replaces  $Z_{ij}$  with the number  $N_{ij} = N_j(X_i)$  of products available in market  $i$  that are owned by firms other than the owner of  $j$ , and one (price as instrument) that replaces  $Z_{ij}$  with  $D_{ij}$ . We demean these instruments so that they have sample mean zero for each product  $j$ . The topmost specification corresponds to the estimated demand model in Miller and Weinberg (2017, Table IV, column (ii)) and features a nest for the inside goods and random coefficients on the constant, number of calories, and price of the product. Subsequent specifications drop elements of the model as indicated. Specifications with a nesting structure include in the instrument set the total number of products in the market, and specifications with a random coefficient on a given attribute include in the instrument set the product of the average consumer income in the market and that attribute, such that all specifications are exactly identified. For each replicate, instrument choice, and specification, we estimate via exactly identified GMM. For specifications with a nesting structure, we exclude any estimate that implies a nesting parameter outside of  $[0, 0.95]$ . The plot depicts the median bias across the simulation replicates, along with (when visible) its 95 percent confidence interval computed based on Mood and Graybill (1963, Sec 16.3).

## A Additional Theoretical Results and Discussion

### A.1 The Importance of Demeaned Instruments in Proposition 1

Assumption 5 plays an important role in Proposition 1. Suppose Assumption 5 fails, but that the selected linear combination of the instruments nonetheless has mean zero on average over  $j$ ,  $\sum_j E [f_j^v (X_i, Z_i)] = 0$ . In this case, define

$$h_+(j) = \frac{E [\max \{f_j^v (X_i, Z_i), 0\}]}{\sum_{j'} E [\max \{f_{j'}^v (X_i, Z_i), 0\}]}, \quad h_-(j) = \frac{E [\max \{-f_j^v (X_i, Z_i), 0\}]}{\sum_{j'} E [\max \{-f_{j'}^v (X_i, Z_i), 0\}]},$$

where  $\sum_j h_+(j) = \sum_j h_-(j) = 1$  by construction. Define  $H_+(x, z|j)$  and  $H_-(x, z|j)$  as in (3), and let  $H_+(j, x, z)$  be the joint distribution on  $\{1, \dots, J\} \times \mathcal{X} \times \mathcal{Z}$  obtained by drawing from  $h_+(j)$  and then from  $H_+(x, z|j)$ . The same argument used to prove Proposition 1 establishes that

$$\int \int E [\tau_i^{R_j} (j_+, x_+, z_+, j_-, x_-, z_-; \tilde{\theta})] dH_+(j_+, x_+, z_+) dH_-(j_-, x_-, z_-) = 0$$

for

$$\tau_i^{R_j} (j_+, x_+, z_+, j_-, x_-, z_-; \tilde{\theta}) \equiv R_{i,j_+} (D_i (x_+, z_+), x_+; \tilde{\theta}) - R_{i,j_-} (D_i (x_-, z_-), x_-; \tilde{\theta}).$$

Hence, in this case  $\tilde{\theta}$  ensures that the average change in the residual from switching the “distribution” of  $(j, x, z)$  from  $H_+$  to  $H_-$  is zero. This is not usually an interesting exercise, however, since the different  $j$  values reflect, e.g., different products in a demand estimation setting, and the difference between the residual for two  $j$  values,  $R_j (Y_i, D_i, X_i; \tilde{\theta}) - R_{j'} (Y_i, D_i, X_i; \tilde{\theta})$ , does not correspond to any causal effect of interest. Correspondingly, when Assumption 5 fails  $\tilde{\theta}$  will be affected by changes in the distribution of potential residuals that do not alter any causal effect, e.g., increasing  $R_j (Y_i (D_i (x, z), x), D_i (x, z), x; \tilde{\theta})$  by one for all  $i$ .

Assumption 5 eliminates this problem by ensuring that  $h_+(j) = h_-(j) = h(j)$  for all  $j$ . Hence, under this assumption the “distribution” on  $j$  is held constant, allowing us to treat  $h(j)$  as a weight and focus solely on the change in the distribution of  $(x, z)$ . Absent Assumption 5 we instead have  $h_+(j) \neq h_-(j)$ , and the moments cannot in general be interpreted in terms of causal effects of  $(x, z)$  regardless of whether the researcher uses included or excluded instruments.

### A.2 Uniqueness of Weighting Scheme in Proposition 1

In this appendix we state conditions under which the weights  $H_+(x, z | j)$ ,  $H_-(x, z | j)$  in Proposition 1 are unique in a particular sense. For simplicity, we assume that the different elements of

the instrument function are linearly independent with finite variance.

**Assumption 8.**  $E [f_j(X_i, Z_i)f_j(X_i, Z_i)']$  has full rank and is finite for all  $j$ .

Let  $F$  be the  $P \times J$  matrix whose  $j$ th column is equal to  $E [f_j(X_i, Z_i)]$ . Denote by  $\mathcal{V}$  the null space of  $F'$ ,  $\mathcal{V} := \{v \in \mathbb{R}^P : F'v = 0\}$ . By definition,  $\mathcal{V}$  is a linear subspace of  $\mathbb{R}^P$  with dimension  $B = P - \text{rk}(F)$ . We associate each  $v \in \mathcal{V} \setminus \{0\}$  to

$$h_v(j) = \frac{E [\max \{v' f_j(X_i, Z_i), 0\}]}{\sum_{j'} E [\max \{v' f_{j'}(X_i, Z_i), 0\}]},$$

$$h_{v,+}(x, z|j) = \frac{\max \{v' f_j(x, z), 0\}}{E [\max \{v' f_j(X_i, Z_i), 0\}]}, h_{v,-}(x, z|j) = \frac{\max \{-v' f_j(x, z), 0\}}{E [\max \{-v' f_j(X_i, Z_i), 0\}]},$$

where by construction  $h_v(j)$ ,  $h_{v,+}(x, z|j)$ , and  $h_{v,-}(x, z|j)$  are non-negative, with  $\sum_j h_v(j) = 1$ , and  $\int h_{v,+}(x, z|j) dG_{X,Z}(x, z) = \int h_{v,-}(x, z|j) dG_{X,Z}(x, z) = 1$ . Hence, we can interpret  $h_v(j)$  as a distribution on  $\{1, \dots, J\}$ , and can regard both  $h_{v,+}(x, z|j)$  and  $h_{v,-}(x, z|j)$  as probability densities on  $\mathcal{X} \times \mathcal{Z}$  with respect to base measure  $G_{X,Z}$ . Denote the corresponding measures by  $H_{v,+}^j(x, z)$  and  $H_{v,-}^j(x, z)$ , respectively. Correspondingly,  $h_v^j(x, z) \equiv h_{v,+}(x, z|j) - h_{v,-}(x, z|j)$  is the density of a signed measure  $H_v^j$  with  $\|H_v^j\|_{TV} = 2$  (for  $\|\cdot\|_{TV}$  the total variation norm) and  $H_v^j(\mathcal{X} \times \mathcal{Z}) = 0$ . Let  $\mathcal{H} = \{(\{H_v^j\}_{j=1}^J, h_v) : v \in \mathcal{V} \setminus \{0\}\}$  denote the set of measures generated by  $\mathcal{V}$ .

Let  $\{v_1, \dots, v_B\}$  be a basis for  $\mathcal{V}$ . When  $B < P$ , let  $\{v_l\}_{l=B+1}^P$  be a basis for the column space of  $F$ . Define  $h_l(j) = h_{v_l}(j)$  and  $h_l^j(x, z) = h_{v_l}^j(x, z)$ . The moment condition may be rewritten as

$$\sum_j \int E [R(Y_i(D_i(x, z), x), D_i(x, z), x; \tilde{\theta}))] \begin{pmatrix} h_1^j(x, z) h_1(j) \\ \vdots \\ h_P^j(x, z) h_P(j) \end{pmatrix} dG_{X,Z}(x, z) = 0. \quad (15)$$

Define  $W_i = (Y_i(\cdot, \cdot), D_i(\cdot, \cdot))$ , and denote its joint distribution by  $G_W \in \mathcal{G}_W$ . Fixing  $\tilde{\theta} \in \Theta$ , define  $\mathcal{G}_W(\tilde{\theta})$  to be the set of joint distributions  $G_W \in \mathcal{G}_W$  that satisfy (15). We impose one assumption:

**Assumption 9.**  $\mathcal{G}_W$  is sufficiently rich, and  $R(y, d, x; \tilde{\theta}) = (R_1(y, d, x; \tilde{\theta}), \dots, R_J(y, d, x; \tilde{\theta}))'$  sufficiently flexible, that for any function  $r : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}^J$ , there exists a joint distribution  $G_W \in \mathcal{G}_W$  such that  $E_{G_W} [R(Y_i(D_i(x, z), x), D_i(x, z), x; \tilde{\theta}))] = r(x, z)$ , where  $E_{G_W}$  is the expectation under  $G_W$ .

Under this assumption, we derive a characterization of the class of weights for which we obtain an interpretation as in Proposition 1.



**Proposition 5.** Consider weights  $\tilde{h}(j) \geq 0$  with  $\sum_j \tilde{h}(j) = 1$ , and probability measures  $\{\tilde{H}_+^j\}_{j=1}^J$  and  $\{\tilde{H}_-^j\}_{j=1}^J$  such that for each  $j$ ,  $\tilde{H}_+^j$  and  $\tilde{H}_-^j$  (i) have disjoint supports and (ii) have densities  $\tilde{h}_+^j$  and  $\tilde{h}_-^j$  with respect to  $G_{X,Z}$  such that  $E[\tilde{h}_+^j(X_i, Z_i)^2]$  and  $E[\tilde{h}_-^j(X_i, Z_i)^2]$  are finite. If Assumption 9 holds, then

$$\sum_j \int \int E_{G_W} \left[ \tau_i^{R_j} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) \right] d\tilde{H}_+^j(x_+, z_+) d\tilde{H}_-^j(x_-, z_-) \tilde{h}(j) = 0 \quad (16)$$

for all  $G_W \in \mathcal{G}_W(\tilde{\theta})$  if and only if for  $\tilde{H}^j = \tilde{H}_+^j - \tilde{H}_-^j$ ,  $(\{\tilde{H}^j\}_{j=1}^J, \tilde{h}) \in \mathcal{H}$ .

**Proof of Proposition 5** First, note that (16) may be rewritten as

$$\sum_j \int E_{G_W} \left[ R_j(Y_i(D_i(x, z), x), D_i(x, z), x; \tilde{\theta}) \right] d\tilde{H}^j(x, z) \tilde{h}(j) = 0.$$

Next, let  $h^{*,j}(x, z) = \left( h_1^j(x, z)h_1(j) \quad \dots \quad h_P^j(x, z)h_P(j) \right)'$ , noting that  $h^{*,j}(x, z) \propto V' f_j(x, z)$  for  $V = [v_1, \dots, v_P]$ . Let  $\tilde{h}^j$  be the density of  $\tilde{H}^j$  with respect to  $G_{X,Z}$ . Under Assumption 9 there exists  $G_W \in \mathcal{G}_W$  such that

$$E_{G_W} \left[ R_j(Y_i(D_i(x, z), x), D_i(x, z), x; \tilde{\theta}) \right] = \tilde{h}^j(x, z) \tilde{h}(j) - \tilde{\beta}' h^{*,j}(x, z),$$

$$\text{for } \tilde{\beta} = E \left[ \sum_j \tilde{h}^j(X_i, Z_i) \tilde{h}(j) h^{*,j}(X_i, Z_i)' \right] E \left[ \sum_j h^{*,j}(X_i, Z_i) h^{*,j}(X_i, Z_i)' \right]^{-1}.$$

By construction,

$$\sum_j \int \left( \tilde{h}^j(x, z) \tilde{h}(j) - \tilde{\beta}' h^{*,j}(x, z) \right) h^{*,j}(x, z) dG_{X,Z}(x, z) = 0,$$

so (15) holds and  $G_W \in \mathcal{G}_W(\tilde{\theta})$ . For this choice of  $G_W \in \mathcal{G}_W(\tilde{\theta})$ , however

$$\begin{aligned} \sum_j \int E_{G_W} \left[ R_j(Y_i(D_i(x, z), x), D_i(x, z), x; \tilde{\theta}) \right] d\tilde{H}^j(x, z) \tilde{h}(j) &= \\ \sum_j E \left[ \left( \tilde{h}^j(X_i, Z_i) \tilde{h}(j) \right)^2 - \tilde{h}^j(X_i, Z_i) \tilde{h}(j) \tilde{\beta}' h^{*,j}(X_i, Z_i) \right] &= \\ \sum_j E \left[ \left( \tilde{h}^j(X_i, Z_i) \tilde{h}(j) - \tilde{\beta}' h^{*,j}(X_i, Z_i) \right)^2 \right]. \end{aligned}$$

Hence, (16) holds only if  $\tilde{h}^j(X_i, Z_i) \tilde{h}(j) = \tilde{\beta}' h^{*,j}(X_i, Z_i)$  almost surely for all  $j$ . Note,

further, that we must have  $V\tilde{\beta} \in \mathcal{V}$ . To see that this is the case, note that  $E[\tilde{h}^j(X_i, Z_i)] \propto E[\tilde{\beta}' h^{*,j}(X_i, Z_i)] \propto \tilde{\beta}' V' E[f_j(X_i, Z_i)]$  so if  $V\tilde{\beta} \notin \mathcal{V}$ , then  $E[\tilde{h}^j(X_i, Z_i)] \neq 0$  for some  $j$ . Since,  $E[\tilde{h}^j(X_i, Z_i)] = \tilde{H}_+^j(\mathcal{X} \times \mathcal{Z}) - \tilde{H}_-^j(\mathcal{X} \times \mathcal{Z})$ , this would imply that at least one of  $\tilde{H}_+^j$  and  $\tilde{H}_-^j$  must not be a probability measure, which is a contradiction. Hence,  $\tilde{h}^j(X_i, Z_i)\tilde{h}(j) = \tilde{\beta}' h^{*,j}(X_i, Z_i)$  for  $V\tilde{\beta} \in \mathcal{V} \setminus \{0\}$ .

It remains to show that  $\left(\left\{\tilde{H}^j\right\}_{j=1}^J, \tilde{h}\right) \in \mathcal{H}$ . By the definition of  $h^{*,j}$ , we know that  $\tilde{h}^j(x, z)\tilde{h}(j) \propto \tilde{\beta}' V' f_j(x, z)$ . Moreover, the assumption of disjoint support for  $\tilde{H}_+^j$  and  $\tilde{H}_-^j$  implies that for  $\|\cdot\|_{TV}$  the total variation norm,  $\|\tilde{H}^j\|_{TV} = 2$ , so  $\sum_j \tilde{h}(j) \|\tilde{H}^j\|_{TV} = 2$ . Note, however, that we also have  $h_{V\tilde{\beta}}^j(x, z)h_{V\tilde{\beta}}(j) \propto \tilde{\beta}' V' f_j(x, z)$ , with  $\sum_j h_{V\tilde{\beta}}(j) \|\tilde{H}_{V\tilde{\beta}}^j\|_{TV} = 2$ . Hence, the constants of proportionality are the same, and  $\tilde{h}^j(x, z)\tilde{h}(j) = h_{V\tilde{\beta}}^j(x, z)h_{V\tilde{\beta}}(j)$ , which along with the constraint on the total variation norm implies that  $\tilde{h}^j(x, z) = h_{V\tilde{\beta}}^j(x, z)$ , from which the result follows. ■

## B Proofs

**Proof of Lemma 1** As discussed in the text following Assumption 2, that assumption implies  $E[\xi_i|X_i, Z_i] = 0$ . Thus, since  $\xi_i = R^*(Y_i, D_i, X_i; \theta_0)$  by Assumption 2(a),  $E[R^*(Y_i, D_i, X_i; \theta_0) f(X_i, Z_i)] = 0$  by the law of iterated expectations, and  $\theta_0$  solves (2). That  $\tilde{\theta} = \theta_0$  is then immediate from Assumption 4, so consistency of  $\hat{\theta}$  for  $\theta_0$  follows from Assumption 3. ■

**Proof of Proposition 1** By Assumption 3 and the fact that  $R_j^*(Y_i, D_i, X_i; \tilde{\theta})$  is scalar,

$$\sum_j E\left[R_j^*(Y_i, D_i, X_i; \tilde{\theta}) v' f_j(X_i, Z_i)\right] = v' \sum_j E\left[R_j^*(Y_i, D_i, X_i; \tilde{\theta}) f_j(X_i, Z_i)\right] = 0.$$

By Assumption 1, however, we can rewrite the term on the left-hand side as

$$\sum_j \int \int E\left[R_{i,j}(D_i(x, z), x; \tilde{\theta})\right] \left(f_{j,+}^v(x, z) - f_{j,-}^v(x, z)\right) dG(x, z),$$

for

$$R_{i,j}(D_i(x, z), x; \theta) = R_j^*(Y_i(D_i(x, z), x), D_i(x, z), x; \theta),$$

$f_{j,+}^v(x, z) = \max\{f_j^v(x, z), 0\}$ , and  $f_{j,-}^v(x, z) = \max\{-f_j^v(x, z), 0\}$ . For each  $j$ , however,

$$\begin{aligned} & \int \int E\left[R_{i,j}(D_i(x, z), x; \tilde{\theta})\right] \left(f_{j,+}^v(x, z) - f_{j,-}^v(x, z)\right) dG(x, z) = \\ & \frac{1}{2} E\left[\left|f_j^v(X_i, Z_i)\right|\right] \int \int E\left[R_{i,j}(D_i(x, z), x; \tilde{\theta})\right] (h_+(x, z|j) - h_-(x, z|j)) dG(x, z) \end{aligned}$$

for  $h_+(x, z|j) = f_{j,+}^v(x, z) / \frac{1}{2}E \left[ \left| f_j^v(X_i, Z_i) \right| \right]$  and  $h_-(x, z|j) = f_{j,-}^v(x, z) / \frac{1}{2}E \left[ \left| f_j^v(X_i, Z_i) \right| \right]$ , where these are both probability densities with respect to  $G(x, z)$  by Assumption 5. The result of the Proposition then follows immediately from the fact that  $h_+(x, z|j)$  and  $h_-(x, z|j)$  are the Radon-Nikodym derivatives of  $H_+(x, z|j)$  and  $H_-(x, z|j)$  with respect to  $G(x, z)$ . ■

**Proof of Corollary 1** This is immediate from Proposition 1 since when  $f_j^v(x, z) = \tilde{f}_j(z)$ , we have  $h_+(x, z|j) = h_+(z|j)$  and  $h_-(x, z|j) = h_-(z|j)$ . ■

**Proof of Corollary 2** Note that with a linear residual

$$\begin{aligned} \tau_i^{R_j^*} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) = \\ \tau_i^{\delta_j} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) - \tilde{\alpha} \tau_i^{D_j} \left( x_+, z_+, x_-, z_-; \tilde{\theta} \right). \end{aligned}$$

Hence, Proposition 1 implies that

$$\begin{aligned} \sum_j \int \int E \left[ \tau_i^{\delta_j} \left( D_{i,j}(x_+, z_+), x_+, D_{i,j}(x_-, z_-), x_-; \tilde{\theta} \right) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j) = \\ \tilde{\alpha} \sum_j \int \int E \left[ \tau_i^{D_j} \left( x_+, z_+, x_-, z_- \right) \right] dH_+(x_+, z_+|j) dH_-(x_-, z_-|j) h(j), \end{aligned}$$

from which the result is immediate. ■

**Proof of Proposition 2** By the fundamental theorem of calculus and the chain rule,

$$\begin{aligned} \tau_i^{\delta_j} \left( D_i(x_+, z_+), x_+, D_i(x_-, z_-), x_-; \tilde{\theta} \right) = \\ \int_0^1 e_j' \left\{ \mathcal{T}_i^{\delta} \left( t, x_+, z_+, x_-, z_-; \tilde{\theta} \right) \tau_i^D \left( x_+, z_+, x_-, z_- \right) + \mathcal{T}_i^{X\delta} \left( t, x_+, z_+, x_-, z_-; \tilde{\theta} \right) \tau_i^X \left( x_+, x_- \right) \right\} dt. \end{aligned}$$

The result is then immediate from Corollary 2. ■

**Proof of Corollary 3** With an excluded instrument we have  $f_j^v(x, z) = \tilde{f}_j(z)$ . Consequently  $h_+(x, z|j) = h_+(z|j)$  and  $h_-(x, z|j) = h_-(z|j)$ , and we can write

$$\begin{aligned} L_X(A_i(\cdot)) = \\ \sum_j \int \int \left( \int_0^1 E \left[ e_j' A_i(t, X_i, z_+, X_i, z_-) \tau_i^X(X_i, X_i) \right] dt \right) dH_+(z_+|j) dH_-(z_-|j) h(j). \end{aligned}$$

Note, however, that  $\tau_i^X(X_i, X_i) = 0$  by definition, so  $L_X(A_i(\cdot)) = 0$  as well. ■

**Proof of Proposition 3** As a first step, note that under the random coefficients logit demand model of Section 5,  $Y_i = Y^M(X_i, \delta_i, \theta)$ , where we denote the researcher's outcome model with  $Y^M(\cdot)$  rather than  $Y^*(\cdot)$  to emphasize that it takes  $\delta_i$  as an argument, not the endogenous variable  $D_i$ . Let  $\mathcal{T}^{\delta Y, M}(X_i, \delta_i; \theta) = \frac{\partial Y_j^M(X_i, \delta_i; \theta)}{\partial \delta}$  denote the model-implied Jacobian of  $Y_i$  with respect to  $\delta_i$  and note that if  $\mathcal{T}^{\delta Y, M}(X_i, \delta_i; \theta)$  is invertible then

$$\frac{\partial \delta(Y_i, X_i; \theta)}{\partial Y_{i,j}} = [\mathcal{T}^{\delta Y, M}(X_i, \delta_i; \theta)^{-1}]_{\cdot, j}$$

by the implicit function theorem. Under the model of Section 5, however,  $\frac{\partial Y_{i,k}^M(X_i, \delta_i; \theta)}{\partial \delta_{i,j}} = \alpha^{-1} \frac{\partial Y_{i,k}^M(D_i, X_i; \theta)}{\partial D_{i,j}}$ , so we can further re-write

$$\frac{\partial \delta(Y_i, X_i; \theta)}{\partial Y_{i,j}} = \alpha [\mathcal{T}^{DY, M}(X_i, \delta_i; \theta)^{-1}]_{\cdot, j},$$

where  $\mathcal{T}^{DY, M}(X_i, \delta_i; \theta)$  is the  $J \times J$  Jacobian matrix of the model-implied shares with respect to the prices,  $[\mathcal{T}^{DY, M}(X_i, \delta_i; \theta)]_{k,j} = \frac{\partial Y_{i,k}^M(X_i, \delta_i; \theta)}{\partial D_{i,j}} \equiv [\mathcal{T}_i^{DY^*}(D_i, X_i; \theta)]_{k,j}$ . As observed in footnote 9, however,

$$\mathcal{T}_i^{D\delta}(d, x) = \frac{\partial}{\partial y} \delta(Y_i(d, x), x; \tilde{\theta}) \mathcal{T}_i^{DY}(d, x).$$

Together with the preceding argument this implies that

$$\mathcal{T}_i^{D\delta}(d, x) = \tilde{\alpha} \mathcal{T}_i^{DY^*}(d, x; \tilde{\theta})^{-1} \mathcal{T}_i^{DY}(d, x).$$

The result is then immediate from Corollary 3. ■

**Proof of Corollary 4** The definition of the residual in the multinomial logit model implies that

$$e'_j \mathcal{T}_i^\delta(D_i, X_i) = \frac{\partial}{\partial d} \log(Y_{i,j}(D_i, X_i)) - \frac{\partial}{\partial d} \log(Y_{i,0}(D_i, X_i)) = \Delta S_{i,j}(D_i, X_i).$$

The result is then immediate from Corollary 3. ■

**Proof of Proposition 4** Using the structure of the residual in Section 6, we can write

$$R_j^*(Y_i, D_i, X_i; \theta) = \delta(Y_{i,j}, X_{i,j}, \theta) - \alpha D_{i,j}.$$

By Assumption 3, we know that

$$\sum_j E \left[ \left( \delta(Y_{i,j}, X_{i,j}, \tilde{\theta}) - \tilde{\alpha} D_{i,j} \right) \tilde{f}_j(Z_{i,j}) \right] = 0.$$

Further, Assumption 7(b) implies that  $Z_{i,j}$  is independent of  $(Y_{i,j}(\cdot), D_{i,j}(\cdot), X_{i,j})$ , so repeating the same arguments used to prove Corollary 3 yields that

$$\tilde{\alpha} = \sum_j \int \int \left( \int_0^1 E \left[ e_j' \mathcal{T}_i^{D\delta} (D_i(t, X_i, z_+, z_-), X_i) \omega_i(X_i, z_+, z_-) \right] dt \right) dH_+(z_+|j) dH_-(z_-|j) h(j)$$

where

$$\omega_i(X_i, z_+, z_-) = \frac{\tau_i^D(X_i, z_+, z_-)}{\sum_j \int \int E \left[ \tau_i^{D_j}(X_i, z_+, z_-) \right] dH_+(z_+|j) dH_-(z_-|j) h(j)}$$

for  $\frac{dH_+}{dG}(z) \propto \tilde{f}_+(z_j)$  and  $\frac{dH_-}{dG}(z) \propto \tilde{f}_-(z_j)$ .

Assumption 7(b) implies that  $H_+(z_+|j)$  and  $H_-(z_-|j)$  differ only in their implied marginal distributions for  $Z_{i,j}$ , while (i) implying the same marginal distributions for  $Z_{i,k}$  with  $k \neq j$  and (ii) implying that  $Z_{i,k} \perp Z_{i,k'}$  for all  $k \neq k'$ . Hence, there exists a joint distribution  $H^*(z_+, z_-|j)$  with marginals  $H_+(z_+|j)$  and  $H_-(z_-|j)$  such that for  $(Z_{i,+}, Z_{i,-}) \sim H^*(z_+, z_-|j)$ ,  $Z_{i,+k} = Z_{i,-k}$  with probability one for all  $k \neq j$ .

Note that we can write

$$\tilde{\alpha} = \sum_j \int \left( \int_0^1 E \left[ e_j' \mathcal{T}_i^{D\delta} (D_i(t, X_i, z_+, z_-), X_i) \omega_i(X_i, z_+, z_-) \right] dt \right) dH^*(z_+, z_-|j) h(j).$$

Assumption 7(a) and the construction of  $H^*$  imply that  $e_k' \omega_i(X_i, Z_{i,+}, Z_{i,-}) = 0$  almost surely for all  $k \neq j$ , which in turn implies that

$$\tilde{\alpha} = \int \left( \int_0^1 E \left[ e_j' \mathcal{T}_i^{D\delta} (D_i(t, X_i, z_+, z_-), X_i) e_j e_j' \omega_i(X_i, z_+, z_-) \right] dt \right) dH^*(z_+, z_-|j).$$

Note, however, that

$$e_j' \mathcal{T}_i^{D\delta} (D_i(t, X_i, z_+, z_-), X_i) e_j = \frac{\partial}{\partial d_j} \delta(Y_{i,j}).$$

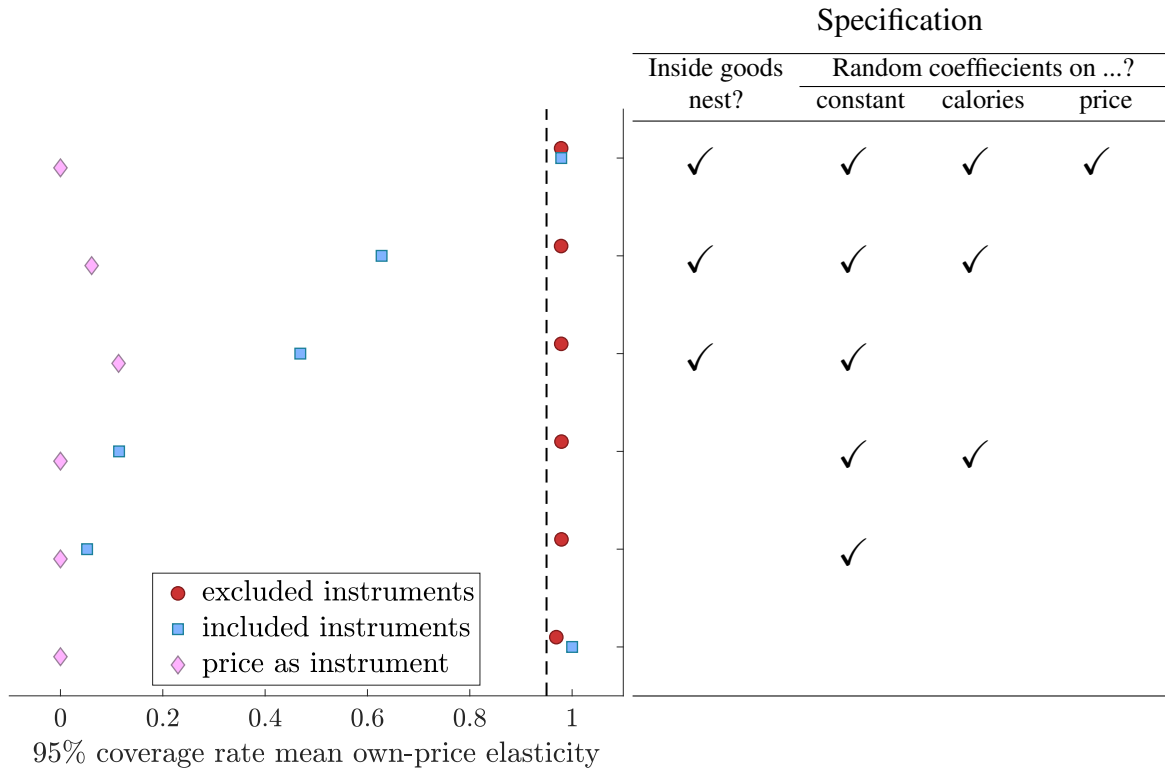
Given the structure of  $\delta(Y_{i,j}, X_{i,j}, \theta)$ , this implies that

$$\begin{aligned} e_j' \mathcal{T}_i^{D\delta} (D_i(t, X_i, z_+, z_-), X_i) e_j &= \frac{\partial}{\partial d_j} \delta(Y_{i,j}(D_i(t, X_i, z_+, z_-), X_i), X_i) = \\ &= \frac{\partial}{\partial d_j} Y_{i,j}(D_i(t, X_i, z_+, z_-), X_i) = \mathcal{T}_{i,j}^{D_j Y} (D_i(t, X_i, z_+, z_-), X_i), \end{aligned}$$

which, noting that  $e_j' \omega_i(X_i, z_+, z_-) = \omega_{i,j}(X_i, z_+, z_-)$ , establishes the result. ■

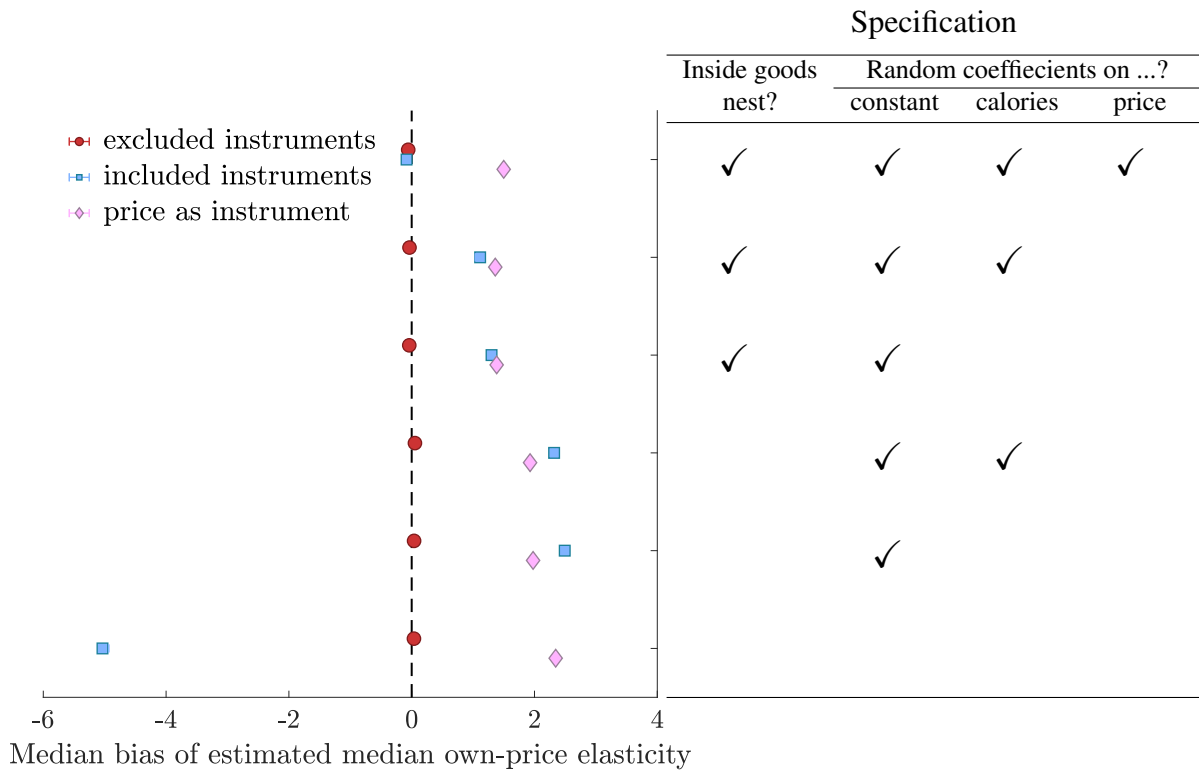
## C Additional Results for the Application to the Demand for Beer

Appendix Figure 1: Coverage rates for confidence intervals on the average own-price elasticity



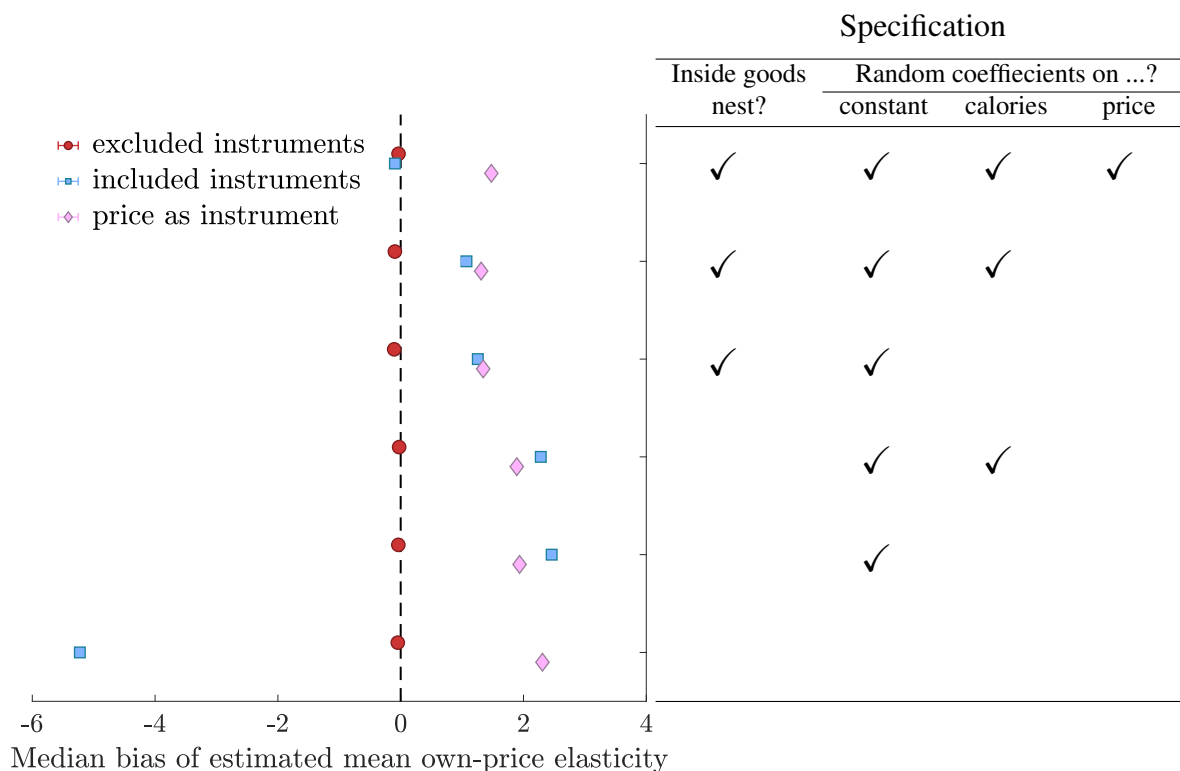
Note: The plot reports the estimated 95 percent coverage rate for confidence intervals on the mean own-price elasticity in the setting of Miller and Weinberg (2017) based on 100 simulations. Each row corresponds to a different demand model specification and each marker shape corresponds to a different choice of instruments. Confidence intervals are constructed via the delta method. The vertical dashed line corresponds to 95 percent coverage. Other details follow Figure 3.

Appendix Figure 2: Estimated median bias for estimators of the median own-price elasticity



Note: The plot reports the estimated median bias for estimators of the median own-price elasticity in the setting of Miller and Weinberg (2017, Table IV) based on 100 simulations. Each row corresponds to a different demand model specification and each marker shape corresponds to a different choice of instruments. Other details follow Figure 3.

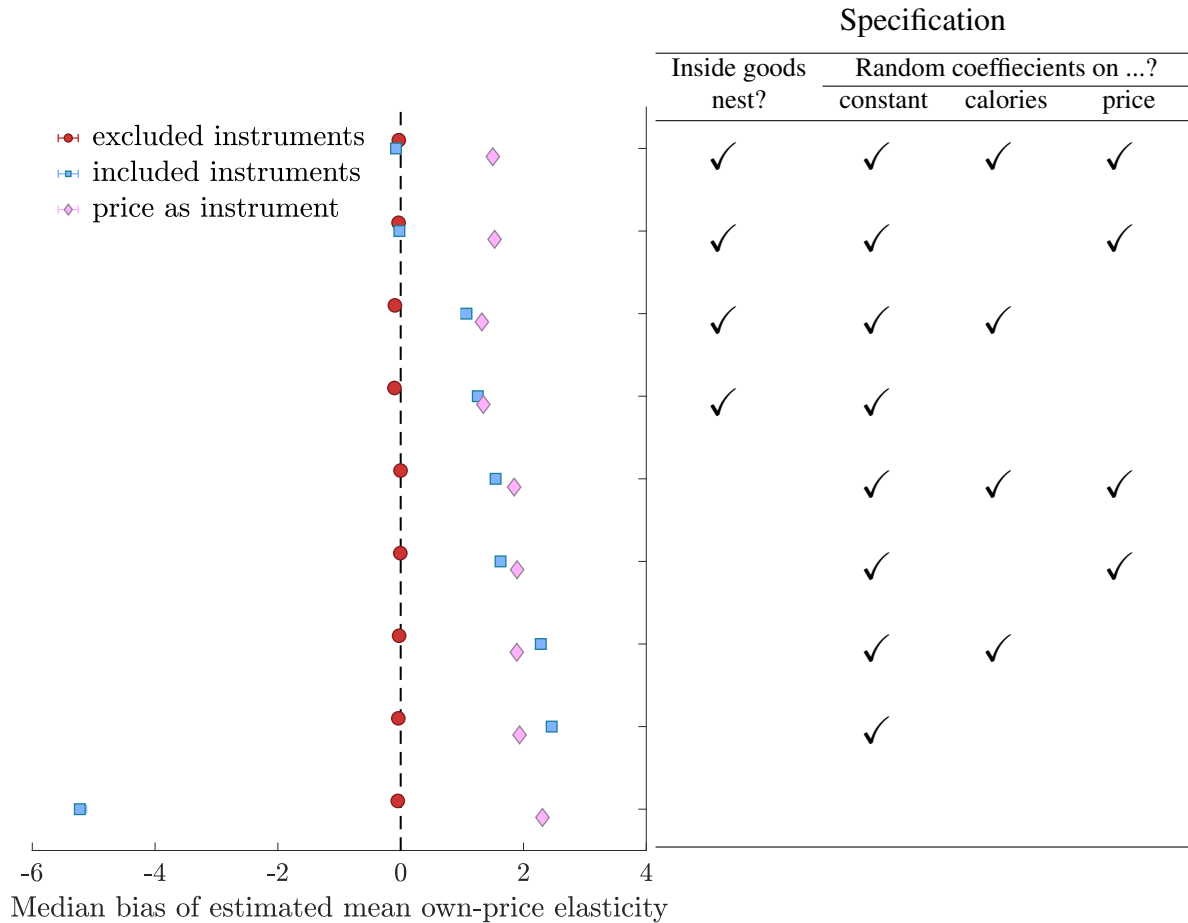
Appendix Figure 3: Estimated median bias for estimators of the average own-price elasticity, all estimates



Note: The plot reports the estimated median bias for estimators of the average own-price elasticity in the setting of Miller and Weinberg (2017) based on 100 simulations. Each row corresponds to a different demand model specification and each marker shape corresponds to a different choice of instruments. For specifications with a nesting structure, we include estimates for which the nesting parameter that imply a nesting parameter outside of  $[0, 0.95]$ . Other details follow Figure 3.



Appendix Figure 4: Estimated median bias for estimators of the average own-price elasticity, including specifications without a linear residual



Note: The plot reports the estimated median bias for estimators of the average own-price elasticity in the setting of Miller and Weinberg (2017) based on 100 simulations. Each row corresponds to a different demand model specification and each marker shape corresponds to a different choice of instruments. Other details follow Figure 3.