# Reinterpreting demand estimation

Jiafeng Chen
Stanford University and SIEPR

Abstract. This paper connects the literature on demand estimation to the literature on causal inference by interpreting nonparametric structural assumptions as restrictions on counterfactual outcomes. It offers nontrivial and *equivalent* restatements of key demand estimation assumptions in the Neyman–Rubin potential outcomes model, for both settings with market-level data (Berry and Haile, 2014) and settings with demographic-specific market shares (Berry and Haile, 2024). This exercise helps bridge the literatures on structural estimation and on causal inference by separating notational and linguistic differences from substantive ones.

*Date*: March 30, 2025. This paper is largely motivated by discussions and issues that arose during a thought-provoking seminar by Steve Berry at the Eddie lunch at Stanford GSB. I thank Isaiah Andrews, Matt Gentzkow, Guido Imbens, Ariel Pakes, Brad Ross, and Jonathan Roth for helpful discussions.

# 1. Introduction

This paper translates key identification results for demand estimation to the Neyman–Rubin potential outcomes model. We show an equivalent formulation, in terms of potential outcomes, of the key conditions for both settings with market-level data (Berry and Haile, 2014) and settings with individual micro-data (Berry and Haile, 2024). In the demand estimation context, potential outcomes encode counterfactual market outcomes, and treatments are conditions of the market (typically prices and characteristics) whose causal effects we would like to learn from existing data.

In our reformulation, identification results in both market- and micro-data settings first assume that the treatment effect of certain market-level interventions $x$ are *latently purgeable*, meaning that there exists some *latent* transformation of outcomes $h$, with $H = h(Y, x)$, such that $x$ has no causal effect on $H$ whatsoever. Second, identification results with market-level data assumes that, in addition, there is some *focal* treatment $w$ which has homogeneous and linear effects on the latently transformed outcome $h(Y, x)$. Identification results with micro-data, on the other hand, do not require such a focal treatment; instead, they assume the latently transformed demographic-specific market shares are almost surely parallel (as functions of the demographic variables $z$). This last assumption is an individualized version of parallel trends assumption in panel data causal inference settings.

Our reformulation is *nontrivial* in the sense that it does not simply declare that the potential outcomes are generated from the corresponding structural model. Instead, the reformulation takes counterfactual outcomes as primitives and makes restrictions in terms of these outcomes directly. The equivalence implies that we can represent the structural errors in the corresponding structural models in terms of these primitives.

These equivalence results are in the spirit of Vytlacil (2002), who shows an equivalence between monotonicity assumptions à la Imbens and Angrist (1994) and structural selection models (Heckman, 1976). They are also in the vein of Angrist *et al.* (2000) applied to state-of-the-art models of markets with multiple goods. Similar to Vytlacil (2002), we show that structural modeling and potential outcomes—in the context of demand estimation—are mainly cultural and notational differences. Substantively, structural assumptions can be equivalently formulated as restrictions on potential outcomes and treatment effects, and structural models may motivate such restrictions (and vice versa).

Our results help bridge the "two cultures," borrowing a phrase from Breiman (2001), of structural econometrics and causal inference. Demand estimation is a foundational exercise in many quantitive structural studies, not only within industrial organization, but more recently in other fields like labor, crime, political economy, and finance as well (Card, 2022; Humphries *et al.*; Longuet-Marx, 2024; Egan *et al.*, 2021). Likewise, the Neyman–Rubin causal inference literature is an influential paradigm across many fields in economics (Goldsmith-Pinkham, 2024; Currie *et al.*, 2020; Conlon and Mortimer, 2021). Our exercise reinforces that these two *languages* have the same capacity of describing economic phenomena. We hope translating between the two is useful for researchers—whose ranks are increasingly numerous—needing fluency in both cultures.

Notwithstanding the equivalence—and just like human languages—the two languages are stylistically different in terms of what each natively thinks of as primitives versus derived quantities. At risk of oversimplification, structural modeling thinks of outcomes as generated by latent and primitive shocks. Potential outcomes models instead treat the counterfactual outcomes as primitives and impose restrictions directly; latent variables are then certain transformations of potential outcomes resulting from the restrictions.

Because of this stylistic difference, structural models appear to be a more native tongue for economic theory, and potential outcomes appear better suited to articulate model misspecification. As a result, we argue that our reformulation sheds light on the restrictions imposed by structural demand models in order to extrapolate counterfactuals, which may lead to informative sensitivity analyses. For the setting with market-level data, we additionally connect our exercise to Andrews *et al.* (forthcoming): We derive conditions under which the nonparametric demand model in Berry and Haile (2014) is misspecified but nevertheless satisfies what Andrews *et al.* (forthcoming) call causally correct specification.

Section 2 contains our results in the setting where the econometrician observes market-level data (Berry and Haile, 2014). Section 3 contains our results in the setting where market shares by demographic subgroup are additionally observed (Berry and Haile, 2024). In each section, we start from a potential outcomes setup. We describe equivalent restrictions on potential outcomes; we then illustrate the identification argument in Berry and Haile (2014, 2024) in terms of the equivalent assumptions. Lastly, we show formally that our restrictions are equivalent to conditions imposed or implied by Berry and Haile (2014, 2024). For settings with market-level data, Section 2.5 additionally connects our formulation to Andrews *et al.* (forthcoming).

## 2. Market-level outcomes

2.1. **Purgeable effects and latent homogeneity.** We begin by considering the potential outcomes model for multiple treatments and a scalar outcome. In a demand context, this corresponds to markets with a single inside option. This is simply to ease exposition: The key elements easily generalize—as we shall—to markets with multiple inside options.

We will refer to observation units as "markets," but the same restrictions on potential outcomes model can apply to other contexts. Consider market counterfactuals drawn from a population $P^*$. A randomly drawn market is associated with potential outcomes $Y(w, x)$. The outcomes respond to a scalar *focal treatment* $w \in \mathcal{W} \subset \mathbb{R}$ and an *auxiliary treatment* $x \in \mathcal{X}$. For concreteness, $w$ represents a special type of characteristics that Berry and Haile (2014) denotes with $x^{(1)}$, and $x$ represents other characteristics and prices.

The potential outcome $Y(w, x)$ encodes what happens to a particular market when the treatments are counterfactually set to some value $(w, x)$. The potential outcomes are random, representing market-level heterogeneity in counterfactual outcomes. We assume that

$$(Y(\cdot, \cdot), W, X) \sim P^*.$$

As is standard, the observed outcome $Y$ is equal to the potential outcome evaluated at the observed values of the treatments $Y = Y(W, X)$. This induces a distribution $P$ over the observed variables:

$$(Y, W, X) \sim P.$$

Suppose that $(W, X)$ are randomly assigned, i.e., $(W, X) \perp\!\!\!\perp Y(\cdot, \cdot)$. Without additional assumptions, we can already identify the average structural function $(w, x) \mapsto \mathbb{E}[Y(w, x)]$, representing counterfactual outcomes averaged over the population of markets. However, it is possible that we would like to predict *individual counterfactual outcomes*—that is, we would like the data to inform potential outcomes $Y(w, x)$ themselves. Section 2.4 in Berry and Haile (2021) makes a strong case that certain policy counterfactuals are really about particular individual markets, and average counterfactual outcomes over many markets—even markets with similar observables—are insufficient.

Formally, identification of the potential outcome $Y(w, x)$ means computing it from the observed outcome $Y$ and the distribution of observable information $P$. Prediction of individual counterfactual outcomes is typically deemed too ambitious in the causal inference literature (since doing so would solve the fundamental problem of causal inference), but it is notably the objective in, e.g., the synthetic control literature.

To do so, we have to entertain stronger restrictions on $Y(w,x)$. The following is one set of restrictions that render $Y(w,x)$ identifiable. Formally, we posit $P^* \in \mathcal{P}^*$, delineated by the following assumptions.[1]

**Assumption 2.1** (Purgeable effect in $x$)**.** *For every $P^* \in \mathcal{P}^*$, there exists some (measurable) function $h(y,x) = h_{P^*}(y,x)$, invertible in $y$, such that $x \mapsto h(Y(w,x),x)$ is constant almost surely: For every $w \in \mathcal{W}$,*

$$h(Y(w,x_1),x_1) - h(Y(w,x_2),x_2) = 0 \text{ for every } x_1, x_2 \in \mathcal{X} \quad P^*\text{-almost surely.}$$

Intuitively, Assumption 2.1 states that the effect of the auxiliary treatment $x$ is purgeable: It can be transformed away with some *unknown* mapping $h$ that preserves information in $Y$. If we define the potential outcome $H(w,x) \equiv h(Y(w,x),x)$, then $x$ has no treatment effect whatsoever—in the sense of Fisher's (1935) sharp null—on $H$. This assumption would trivially hold if one could define $h$ that destroys variation in $Y$. For this reason, we require $h$ to be invertible in $y$.

The next assumption additionally imposes that the treatment effect of $w$ on $H(w) = H(w,x)$ is constant, linear, and nonzero.

**Assumption 2.2** (Latent constant linear treatment effects in $w$)**.** *For each member of $\mathcal{P}^*$, for $h$ that satisfy Assumption 2.1, the effect of $w$ on $h(Y(w,x),x)$ is constant, linear, and nonzero in $w$: There exists some $b_{P^*} \neq 0$ such that, for all $x \in \mathcal{X}, w_1, w_0 \in \mathcal{W}$,*

$$h(Y(w_1,x),x) - h(Y(w_0,x),x) = b_{P^*}(w_1 - w_0) \quad P^*\text{-almost surely.}$$

*By redefining $h$ if necessary, we can choose $b_{P^*} = 1$.*

Constant treatment effects are frequently criticized as a strong assumption in the causal inference literature (e.g., in Chen and Roth, 2024; Mogstad and Torgovitsky, 2024). That said, such restrictions are also routinely imposed in applied work (e.g. Abdulkadıroğlu *et al.*, 2022).[2] Since the function $h$ is unknown and allowed to vary with $P^*$, Assumption 2.2 is a weaker assumption than constant (linear) treatment effects.

[1] For simplicity, assume that all members of $\mathcal{P}^*$ induce distributions $P$ that share the same support and are mutually absolutely continuous with each other.

[2] The results of Blandhol *et al.* (2022) suggest that the causal interpretation of two-stage least-squares with covariates is often difficult absent imposing homogeneous treatment effects. As a result, whether or not intended, empirical work using these methods in such settings implicitly impose effect homogeneity.

Moreover, if the goal is to identify individual counterfactual outcomes $Y(w, x)$, constant treatment effects assumptions are unavoidable. If we could point-identify unit-level counterfactual outcomes, then there necessarily is a mapping $\mathfrak{C}$—exactly what we get out of a point-identification argument—such that the counterfactual outcome is deterministically related to the observed data $(Y, W, X)$ through $\mathfrak{C}$:

$$Y(w, x) = \mathfrak{C}(w, x; (Y, W, X); P) \quad P^*\text{-almost surely.}$$

One could view the above display as a general form of effect homogeneity, since the mapping $\mathfrak{C}$ does not depend on unit-level unobservables $U$, and deterministically links the observed outcome $Y$ to the counterfactual outcomes $Y(w, x)$. Thus, effect homogeneity is intrinsic to forecasting individual-level counterfactual outcomes, and Assumption 2.2 puts some additional structure beyond this minimal requirement.

Assumptions 2.1 and 2.2 can be strengthened so that the distinction between $w$ and $x$ disappears: We may be willing to impose that effects are constant and linear in *both* $(w, x)$ for a latent transformation, assuming $\mathcal{X} \subset \mathbb{R}^k$:

$$g_{P^*}(Y(w_1, x_1)) - g_{P^*}(Y(w_2, x_2)) = b_{P^*}(w_1 - w_2) + c'_{P^*}(x_1 - x_2) \quad b_{P^*} \neq 0.$$

In this case, Assumptions 2.1 and 2.2 are satisfied with the choice $h(y, x) = g_{P^*}(y) - c'_{P^*}x$.

Under Assumptions 2.1 and 2.2, we can write $Y(w, x)$ in the following form:

$$h(Y(w_1, x_1), x_1) - h(Y(w_0, x_0), x_0) = h(Y(w, x_0), x_0) - h(Y(w_0, x_0), x_0) \quad \text{(Assumption 2.1)}$$

$$= w - w_0. \tag{1}$$

$$\implies Y(w, x) = h^{-1}\left(w + h(Y(W, X), X) - W, x\right) \tag{2}$$

Thus, if $h$ is identifiable from $P$ up to a level shift, then we can extrapolate to all potential outcomes $Y(w, x)$ from the observed outcome $Y(W, X)$.

2.2. **Identification.** The identification of $h$ depends on assumptions on treatment assignment or availability of randomly assigned instruments. The demand estimation literature often imposes the following treatment/design-based assumption

**Assumption 2.3.** *(1) The focal treatment $W$ is randomly assigned $W \perp\!\!\!\perp Y(\cdot, \cdot)$.*

*(2) There is a set of randomly assigned instruments $Z$ for the auxiliary treatments $X$, which may include members of $X$, and $(W, Z) \perp\!\!\!\perp Y(\cdot, \cdot)$.*

Finally, the instruments are assumed to satisfy a completeness assumption (Newey and Powell, 2003), which says that the instruments $(W, Z)$, in a sense, capture all relevant variation in $(Y, X)$. In causal inference, completeness assumptions are invoked in proximal inference (Tchetgen Tchetgen *et al.*, 2024) and bridge function literatures (Imbens *et al.*, 2024).

**Assumption 2.4.** *The joint distribution $P$ of $(Y, W, X, Z)$ is complete in the following sense: No nonzero, integrable function $g(y, x)$ satisfies $\mathbb{E}_P[g(Y, X) \mid W, Z] = 0$. The function $h$ in Assumption 2.1 is assumed to be integrable.*

**Proposition 2.1.** *Under Assumptions 2.1 to 2.4, $h$ is identified up to a location shift. By (2), $Y(w, x)$ is identified from $((Y, W, X), P)$.*

*Proof.* Fix some $w_0, x_0$, Assumptions 2.1 and 2.2 imply that

$$h(Y, X) = (W - w_0) + h(Y(w_0, x_0), x_0).$$

Take conditional expectations under $W, Z$:

$$\begin{aligned} \mathbb{E}[h(Y, X) \mid W, Z] &= (W - w_0) + \mathbb{E}\left[h(Y(w_0, x_0), x_0) \mid W, Z\right] \\ &= (W - w_0) + C_{P^*} \qquad \text{(Assumption 2.3)} \end{aligned}$$

for some constant $C_{P^*} = \mathbb{E}[h(Y(w_0, x_0), x_0)]$. Consider the solutions $f$ to the integral equation

$$\mathbb{E}[f(Y, X) \mid W, Z] = W - w_0$$

Note that $f(y, x) = h(y, x) - C_{P^*}$ is a solution. By Assumption 2.4, it is the unique solution. Thus $h(\cdot, \cdot)$ is identified up to a location shift. $\square$

2.3. **Equivalence in markets with one good.** We follow the exposition in Section 5 of Berry and Haile (2021) and specialize to the case with a single inside good (in their notation, $J = 1$). We will again partition the characteristics and the prices of the good into $W$ and $X$, where $W$ is a special scalar characteristic whose existence is implied by one of the assumptions in Berry and Haile (2014). $X = (D, X_2)$ contains other characteristics as well as prices $D$.[3] Berry and Haile (2021) represent market shares $Y$ as structural equations

$$Y = \mathfrak{s}(W, D, X_2, \tilde{\xi}) \in (0, 1).$$

For identification, Berry and Haile (2014) impose economic assumptions that imply the following conditions. We impose these as assumptions since they are key ingredients in the identification argument, as outlined by Berry and Haile (2021).

[3] Berry and Haile (2021) use $x^{(1)}$ to denote $W$, $x^{(2)}$ to denote $X_2$, and $p$ to denote $D$.

**Assumption 2.5** (Linear index, Assumption 5.1 of Berry and Haile (2021))**.** *For some $(W, \xi)$ such that $\delta = W + \xi$ and $\sigma(\cdot)$, we can reparametrize the structural equations*

$$\mathfrak{s}(W, D, X_2, \tilde{\xi}) = \sigma(\delta, D, X_2).$$

**Assumption 2.6** (Invertible demand (5.12), Berry and Haile (2021))**.** *The function $\sigma$ is invertible:*

$$\delta = \sigma^{-1}(Y, D, X_2).$$

**Theorem 2.2.** *Assumptions 2.5 and 2.6 are equivalent to Assumptions 2.1 and 2.2.*

This equivalence result shows that the key conditions imposed in Berry and Haile (2014) can be reformulated as Assumptions 2.1 and 2.2. Subjected to these assumptions, with the additional assumptions of instruments and completeness, identification of $\sigma$ and $\xi$—equivalent to the identification of $Y(w, x)$—is obtained through Proposition 2.1.

*Proof.* We first show that Assumptions 2.5 and 2.6 imply Assumptions 2.1 and 2.2. It suffices to pick some $h$. By Assumption 2.6, we can choose $h(y, x) = \sigma^{-1}(y, d, x_2)$ since $X = (D, X_2)$. This function is invertible by assumption. Moreover,

$$h(Y(w, x), x) - h(Y(w', x'), x') = w - w'$$

by Assumption 2.5. This proves both Assumptions 2.1 and 2.2 by setting $(w, w', x, x')$ appropriately.

For the other direction, under Assumptions 2.1 and 2.2, by (2), we know that for any choice $(w_0, x_0)$ in the support,

$$h(\mathfrak{s}(w, x, \tilde{\xi}), x) = w \underbrace{-w_0 + h(\mathfrak{s}(w_0, x_0, \tilde{\xi}), x_0)}_{\xi}.$$

For this choice of $\xi$, we have that almost surely $\mathfrak{s}(w, x, \tilde{\xi}) = h^{-1}(w + \xi, x)$. Thus Assumption 2.5 is satisfied by choosing $\sigma(\delta, d, x_2) = h^{-1}(\delta, (d, x_2))$. The assumption that $h$ is invertible in (2) implies Assumption 2.6. $\square$

2.4. **Markets with multiple inside options.** To generalize to vector-valued outcomes, we next state the $J$-dimensional analogue of Assumptions 2.1 and 2.2. Here, the focal treatment would also be $J$ dimensional ($\mathcal{W} \subset \mathbb{R}^J$).

**Assumption 2.7** (Purgeable effects and latent constant linear effects)**.** *For each member of $\mathcal{P}^*$, there exists some invertible function $h = h^*_P$ and invertible $J \times J$ matrix*

$B_{P^*}$, *normalized to* $I_J$, *such that for all* $w_1, w_2, x_1, x_2$,

$$h_{P^*}(Y(w_1, x_1), x_1) - h_{P^*}(Y(w_2, x_2), x_2) = w_1 - w_2.$$

Assumption 2.7 is the $J$-dimensional analogue to Assumptions 2.1 and 2.2. Since we restrict effects to be linear, it is reasonable to consider $J$ focal treatments in order to capture the effect on $J$-dimensional outcomes.

By an analogous argument, this assumption is equivalent to the key conditions in Berry and Haile (2014) for the general case with $J$ inside goods. We state the result for completeness. Now, let $W = (W_1, \ldots, W_J) \in \mathbb{R}^J$ denote a special characteristic, one for each good. Let $X = (D_j, X_j)_{j=1}^J$ denote the other characteristics and prices, and assume the structural model for the market shares of each of the $J$ inside options

$$Y = (Y_1, \ldots, Y_J) = \mathfrak{s}(W, X, \tilde{\xi}).$$

**Assumption 2.8** (Linear index)**.** *For some* $\delta = (\delta_1, \ldots, \delta_J) = W + \xi$, *we can write*

$$\mathfrak{s}(W, X, \tilde{\xi}) = \sigma(\delta, D, X_2).$$

**Assumption 2.9** (Invertibility)**.** *The function* $\sigma$ *admits an inverse where*

$$\delta = \sigma^{-1}(X, Y, D).$$

Completely analogous to Theorem 2.2, we have the following theorem.

**Theorem 2.3.** *Assumptions 2.8 and 2.9 are equivalent to Assumption 2.7.*

2.5. **Misspecification and identification.** The formulation Assumptions 2.1 and 2.2 allows close connections to the literature on misspecification of structural models. Potential outcomes models are natural for studying misspecification, as they do not presume the existence of special latent variables and take the counterfactual outcomes themselves as model primitives. The potential outcomes model—possibly before imposing Assumptions 2.1 and 2.2—is what Andrews *et al.* (forthcoming) term a nesting model. This subsection details the connection to their results. In particular, we derive conditions under which Assumptions 2.1 and 2.2 may be misspecified yet correctly capture the right counterfactual outcomes in a particular treatment (e.g., prices). A byproduct of this analysis is an interesting tradeoff between model flexibility—so as to guard against misspecification—and identification under misspecification.

Consider a nesting model $\mathcal{P}^{**}$, which may not obey Assumptions 2.1 and 2.2, from which $P^{**}$ is chosen by Nature. Suppose we can partition $X = (D, X_2)$ for some treatment of interest $D$ (typically price). Andrews *et al.* (forthcoming) define a notion called *causally correct specification* of a structural model $\mathcal{P}^*$. Loosely speaking, a structural model $\mathcal{P}^* \subset \mathcal{P}^{**}$ is causally correctly specified if, for every $P^{**}$, there is *some* member $P^* \in \mathcal{P}^*$ under which all partial derivatives of $\frac{\partial}{\partial d} Y(w, d, x_2)$ under $P^{**}$ are reproduced by corresponding partial derivatives under $P^*$.[4] The following definition makes this precise, defining causal correct specification using the representation in Proposition 2 in Andrews *et al.* (forthcoming) directly.

**Definition 2.4** (Causally correct specification)**.** Consider a structural model $\mathcal{P}^*$ under which potential outcomes take the form

$$Y_{\text{model}}(w, d, x_2) = Y_{P^*}(w, d, x_2, \xi) \;\; P^*\text{-almost surely}$$

for $\xi \in \mathbb{R}^J$. We say that $\mathcal{P}^*$ *is causally correctly specified with respect to $d$ for $\mathcal{P}^{**}$* if, for every member $P^{**} \in \mathcal{P}^{**}$, the potential outcomes have the following representation: For some $P^* \in \mathcal{P}^*$, some function $L$ and some choices of unobservable random variables $(\xi, V)$,

$$Y(w, d, x) = Y_{P^*}(w, d, x, \xi + L(w, x, V)) \quad P^{**}\text{-almost surely}$$

where the choice of $\xi$ is such that $Y_{P^*}(w, d, x, \xi)$ is distributed according to $P^*$.

When $\mathcal{P}^*$ satisfies Definition 2.4, for some member $P^*$, its model-implied partial derivative in $d$ matches all corresponding partial derivatives of $Y(w, d, x)$, assuming both sets of derivatives exist.

A natural question is when $\mathcal{P}^*$ defined by Assumptions 2.1 and 2.2 is causally correctly specified. The answer is when Nature's model $\mathcal{P}^{**}$ satisfies a weaker version of purgeable effects (Assumption 2.1), where we only require effects be purgeable in $d$.

**Assumption 2.10** (Purgeable effects in $d$)**.** *For all $P^{**} \in \mathcal{P}^{**}$, there exists some measurable $h^{**}(y, d, x_2)$ (which does not need to depend on $x_2$), invertible in $y$, such that, for all $w, x_2, d_1, d_2$ in their respective supports,*

$$h^{**}(Y(w, d_1, x_2), d_1, x_2) - h^{**}(Y(w, d_2, x_2), d_2, x_2) = 0 \quad P^{**}\text{-almost surely.}$$

**Proposition 2.5.** *Let $\mathcal{P}^*$ be defined by Assumptions 2.1 and 2.2.*[5] *Then $\mathcal{P}^*$ is causally correctly specified with respect to $d$ for $\mathcal{P}^{**}$ satisfying Assumption 2.10.*

[4]This assumption simply imposes that a given structural model $\mathcal{P}^*$ is rich enough to model the causal effect of $d$ correctly, yet $P^*$ may not equal $P^{**}$, nor is it necessarily identified from the data.

[5]This means that for any invertible $h$, there exists some member $P^* \in \mathcal{P}^*$ that obeys Assumptions 2.1 and 2.2 with respect to $h$.

*Proof.* Given $P^{**}$ and $h^{**}$ by Assumption 2.10, choose the member of $\mathcal{P}^*$ under which $h = h^{**}$. Fix some $x_0, w_0$, the potential outcomes under the model $\mathcal{P}^*$ can be represented as

$$Y_{\text{model}}(w, x) = h^{-1}\left(w + \underbrace{h(Y(w_0, x_0), x_0) - w_0}_{\xi}, x\right) \equiv Y_h(w, x, \xi).$$

Under $\mathcal{P}^{**}$,

$$\begin{aligned} &h(Y(w, d, x_2), d, x_2) - h(Y(w_0, d_0, x_{20}), d_0, x_{20}) \\ &= h(Y(w, d_0, x_2), d_0, x_2) - h(Y(w_0, d_0, x_{20}), d_0, x_{20}) \qquad \text{(Assumption 2.10)} \end{aligned}$$

Define the last line on the display as

$$L(w, x_2, \underbrace{Y(\cdot, d_0, \cdot)}_{V}) - w_0.$$

Thus,

$$h(Y(w, d, x_2), d, x_2) = \xi + L(w, x_2, V) \implies Y(w, d, x_2) = Y_h(w, x, \xi + L(w, x_2, V)).$$

This concludes the proof. $\square$

**Remark 2.6** (Tradeoff in misspecification-robustness and identification)**.** This analysis highlights an interesting tradeoff between flexible parametrization—making causal correctness more plausible—and identification. Under models that impose Assumptions 2.1 and 2.2, researchers typically exploit the moment condition (e.g., in Proposition 2.1)

$$\mathbb{E}\left[h(Y, D, X_2) - W \mid Z, W\right] \equiv \mathbb{E}[\xi \mid Z, W] = 0$$

for estimation.[6] Equivalently, for any set of test functions $T(Z, W)$, we may exploit the moment condition $\mathbb{E}[(h(Y, D, X_2) - W)T(Z, W)] = 0$ to estimate $h(\cdot)$.

Because of the possibility that, even assuming causally correct specification, there is a misspecified component $L(W, X_2, V)$,

$$\mathbb{E}\left[h(Y, D, X_2) - W \mid Z, W\right] = \mathbb{E}[L(W, X_2, V) \mid Z, W] \neq 0,$$

Andrews *et al.* (forthcoming) recommend recentering instruments (Borusyak and Hull, 2023) to purge them from accidentally loading on $(W, X_2)$. Namely, restricting to $\tilde{T}(Z, W)$ such that $\mathbb{E}[\tilde{T}(Z, W) \mid W, X_2] = 0$, Andrews *et al.* (forthcoming) recommends

[6]We can normalize $\mathbb{E}[\xi] = 0$ since we only need to estimate $h$ up to a location shift.

exploiting the following moment conditions, valid if $Z$ is an external instrument (assuming $Z \perp\!\!\!\perp (V, Y(\cdot,\cdot)) \mid X$):

$$\mathbb{E}[(h(Y,D,X_2) - W)\tilde{T}(Z,W)] = \mathbb{E}[\mathbb{E}[L(W,X_2,V) \mid W, X_2] \cdot \tilde{T}(Z,W)] = 0.$$

In parametric demand models (i.e., $h = h(\cdot;\theta)$ for some Euclidean $\theta$), the variation in $\tilde{T}(Z,W)$ orthogonal to $(W, X_2)$ is potentially sufficient for identifying the parameter $h$. However, for nonparametric demand models, even though the model is causally correctly specified, the parameter $h$ that reproduces all causal summaries in $d$ may not be pinned down by the data, because variation in $(W, X_2)$ may be necessary to identify $h$ but are contaminated by $L(W, X_2)$.[7] When we lack identification, Deaner (forthcoming) shows that, unfortunately, even if the misspecification $\mathbb{E}[L(W,X_2,V) \mid Z, W]$ is known to be small, the ill-posedness of nonparametric instrumental variables can nonetheless result in large identified sets for $h$. ■

## 3. Markets with micro-data

**3.1. Latent purgeability and latent parallel trends.** For micro-data settings (Berry and Haile, 2024), consider a set of market-level interventions $x$ (e.g., prices). Here, let $w$ denote observable demographics. The potential outcome is a random *function* mapping demographics to their specific market shares:

$$Y(x) : \mathcal{W} \to (0,1).$$

We likewise keep the exposition with a single inside option. Extension to $J$ options is straightforward. We denote evaluation of this potential outcome at a value $w$ as $Y(x)[w]$, which is the market share of demographic $w$ for a randomly drawn market—were the market to receive intervention $x$ exogenously. Berry and Haile (2024) term this object the conditional demand system.[8]

[7]To see this, suppose

$$\mathbb{E}_P\left[h(Y,D,X_2) - W - L(W,X_2,V) \mid Z, W\right] = 0 \tag{3}$$

is the only restriction on $h$ for an unknown $L$ and unknown $V$. Suppose there exists some choice of $L_0, V$ and $\tilde{h}(y,d,x_2)$, where $\tilde{h}$ depends nontrivially on $d$, such that

$$\mathbb{E}[\tilde{h}(Y,D,X_2) \mid Z, W] = \mathbb{E}[L_0(W,X_2,V) \mid Z, W].$$

Then for any given $h_0$, $(h, L) = (h_0, 0)$ and $(h, L) = (h_0 + \tilde{h}, L_0)$ are observationally equivalent in the sense that they both satisfy (3). Appendix A.1 shows an example in which such a choice of $(\tilde{h}, L_0)$ can be made.

[8]Unfortunately, to keep notation unified with the previous section, this notation differs from Berry and Haile (2024). What we call $x$ is meant to capture what Berry and Haile (2024) denote as price $p$. What we call $w$ is $z$ in Berry and Haile (2024). Finally, Berry and Haile (2024) additionally has a market-level variable $X_t$ that denotes other market interventions or characteristics that may or may not be exogenous. This variable is essentially conditioned upon throughout their identification

Compared to $Y(x, w)$, this notation captures the fact that $Y(x)[w_1] - Y(x)[w_0]$ are not causal comparisons in $w$.[9] This formulation has some similarities to panel data (as pointed out in Berry and Haile, 2021), since we may view panel data as a setting in which $\mathcal{W} = \{1, \ldots, T\}$ is the time indices, and $Y(x)[t]$ as the time-$t$ potential outcome.

As before, the latent population is $(Y(\cdot)[\cdot], X) \sim P^* \in \mathcal{P}^*$, from which we observe $(Y[\cdot], X) \sim P$ where $Y = Y(X)$.

Here, we impose the following assumptions: The first is an analogue of Assumption 2.1.

**Assumption 3.1** (Latent purgeable effects, micro-data)**.** *For all $P^* \in \mathcal{P}^*$, there exists a measurable function $h = h_{P^*}(y, x)$, invertible in $y$ such that for all $w \in \mathcal{W}$ and all $x_1, x_2 \in \mathcal{X}$,*

$$h(Y(x_1)[w], x_1) - h(Y(x_2)[w], x_2) = 0 \quad P^*\text{-almost surely.}$$

The second replaces Assumption 2.2 with a parallel-trends analogue.[10]

**Assumption 3.2** (Latent individual parallel trends)**.** *There exists a fixed $w_0 \in \mathcal{W}$ an invertible and differentiable function $g(w) = g_{P^*}(w)$ with $g'(w_0) \neq 0$ such that the map $h$ in Assumption 3.1 additionally satisfies, for all $x \in \mathcal{X}$ and $w \in \mathcal{W}$*

$$h(Y(x)[w], x) - h(Y(x)[w_0], x) = g(w) - g(w_0) \quad P^*\text{-almost surely.}$$

*Redefining $h$ if necessary, we may normalize $g'(w_0) = 1$ and $g(w_0) = 0$.*

Instead of assuming that there is some focal treatment whose effect is homogeneous and linear, Assumption 3.2 imposes that the random function

$$w \mapsto h(Y(x)[w], x) \tag{4}$$

has parallel sample paths equal to $g(w) + h(Y(x_0)[w_0], x_0)$. When we think of $w$ as a time period, and of $H_w(x) = h(Y(x)[w], x)$ as a transformed potential outcome path, this assumption is an individual version of parallel trends on $H_w(x)$. It imposes that differences $H_{w_1}(x) - H_{w_2}(x)$ are not only mean independent of other variables—which

argument for what they term the "conditional demand system", which considers counterfactuals in which $X_t$ is fixed at the observed values. We suppress $X_t$ in our exposition. If there are no $X_t$, this setup corresponds to the demand system in Berry and Haile's (2024) terminology.

[9] Consider the following example: In the Cambridge, MA tourism market in July 2026, those who have stayed in the Royal Sonesta in July 2025 ($w$) are very likely to choose the Royal Sonesta again in July 2026 ($Y(x)[w]$ is high), because they are likely attendees to the NBER Summer Institute meetings. However, had we exogenously assigned a random individual visiting Cambridge, MA in 2025 to the Royal Sonesta, she would not return with equal propensity.

[10] The connection to difference-in-differences codifies some existing intuition. Steve Berry also made references to difference-in-differences for this argument during his seminar at Stanford in March 2025, despite the terminology not featuring in Berry and Haile (2024).

is imposed by typical parallel trends—but are identical and nonrandom. Despite this strong parallel trends assumption, positing that it holds under an unknown transformation $h(\cdot, x)$ does provide considerable flexibility.

The $J$ dimensional analogue of Assumptions 3.1 and 3.2 simply imposes that $h, g, w$ are all $J$-dimensional and normalizes the Jacobian $\frac{d}{dw}g(w_0) = I_J$.

3.2. **Identification under latent purgeability and latent parallel trends.** As before, Assumptions 3.1 and 3.2 allows us to write

$$Y(x)[w] = h^{-1}\left(h(Y(X)[w_0], X) + g(w), x\right) \tag{5}$$

Upon identification of $h(\cdot, \cdot)$ (up to a level shift) and $g(\cdot)$, we would be able to predict $Y(x)[w]$ at counterfactual $x$ values.

Assumptions 3.1 and 3.2 are in fact sufficiently restrictive to allow for identification of $g(\cdot)$ and of $h(\cdot, x)$, just by analyzing the joint distribution $(Y[\cdot], X) \sim P$—after imposing a few regularity assumptions. Notably, this identification argument relies solely on the outcome model imposed, and does not rely on assumptions about the random assignment of $X$ or the availability of randomly assigned instruments. To show this identification, subjected to our equivalence result shortly, one could follow Berry and Haile's (2024) argument in their Lemma 2, Lemma 3, and Corollary 1, as well as Berry and Haile (2021) Section 7.2 for a simplified presentation.

We briefly present an informal version of this argument that includes the key insight, possibly imposing stronger regularity assumptions. Note that Assumptions 3.1 and 3.2 impose that for all $x$,

$$g(w) - g(w') = h(Y(x)[w], x) - h(Y(x)[w'], x).$$

Since both $g$ and $h$ are invertible, $w \neq w'$ implies $Y(x)[w] \neq Y(x)[w']$. Hence, we can consider the *inverse* potential outcomes map $W(x)[y]$. Observing the distribution of $Y(X)[\cdot]$ is equivalent to observing the inverse $W(X)[\cdot]$. Now, suppose (i) $y \mapsto W(x)[y]$ is continuously differentiable for every $y$, (ii) $g, g^{-1}$ are uniformly continuous and continuously differentiable, and (iii) $h(y, X)$ is differentiable in $y$ with a nonzero partial derivative for almost all $y$ and almost every $X$.

Then under Assumptions 3.1 and 3.2, for a fixed $y$, the distribution of $W(X)[y]$ is known given $P$. Moreover, this distribution relates to $g^{-1}$ and $h$ by

$$\begin{aligned} W(X)[y] &= g^{-1}(\underbrace{h(Y(X)[W(X)[y]], X)}_{h(y,X)} - h(Y(X)[w_0], X)) \\ &= g^{-1}(h(y, X) - h(Y[w_0], X)). \end{aligned}$$

Differentiating both sides in $y$,

$$W(X)'[y] = \frac{1}{g'\left(g^{-1}(h(y,X) - h(Y[w_0],X))\right)} \frac{\partial h(y,X)}{\partial y} = \frac{1}{g'(W(X)[y])} \frac{\partial h(y,X)}{\partial y} \quad (6)$$

Condition on a value of $X$ and consider the distribution $W(X)[y] \mid X$. Fix $y$ at which $\frac{\partial h(y,X)}{\partial y}$ is nonzero. At this value of $y$, for values $w_1, w_2$ in the support of $W(X)[y] \mid X$ with corresponding values $w_1', w_2'$ of $W(X)'[y]$, then

$$\frac{g'(w_1)}{g'(w_2)} = \frac{w_2'}{w_1'} \quad (7)$$

is identified, since the right-hand side is known. This argument means that the ratio in $g'(\cdot)$ is identified for all pairs $w_1, w_2$ that are in the support of $W(X)[y] \mid X$ for some values of $X$ and $y$.

As long as we can connect an arbitrary value $w$ to the normalized value $w_0$ through a chain of such pairs $((w, w_1) \to (w_1, w_2) \to \cdots \to (w_{n-1}, w_n) \to (w_n, w_0))$, (7) allows us to identify $g(w)$, since we normalized $g'(w_0) = 0$ and $g(w_0) = 1$. Identifying $g(\cdot)$ allows us to identify $h_y(y,x) \equiv \frac{\partial h(y,x)}{\partial y}$ for $x$ values in the support of $X$ by (6). Integrating $h_y(y,x)$ in $y$ identifies differences $h(y_1,x) - h(y_2,x)$. We state these identification results as high-level assumptions directly, but they can be obtained by imposing lower-level assumptions on top of Assumptions 2.1 and 3.2.

**Assumption 3.3** (Outcome-modeling identification)**.** *Assumptions 2.1 and 3.2 hold. The function $g(\cdot)$ in Assumption 3.2 is known. The function $h$ in Assumption 2.1 has known dependence on $y$: That is, for all $y, y_1, x$, $\frac{\partial}{\partial y} h(y,x)$ is known, and $h(y_1,x) - h(y,x)$ is known. The functions $h, g$ are integrable.*

The final setup of the identification—solely to identify baseline values $x \mapsto h(y_0, x)$—again relies on randomly assigned instruments which are sufficiently strong in the sense of Newey and Powell (2003).

**Assumption 3.4.** *There is some randomly assigned instrument $Z \perp\!\!\!\perp Y(\cdot)[\cdot]$. It satisfies a completeness assumption for all $P$: No nonzero integrable function $q(X)$ satisfies $\mathbb{E}_P[q(X) \mid Z] = 0$.*

**Proposition 3.1** (Berry and Haile (2024), Lemma 4 and Theorem 1)**.** *Under Assumptions 3.3 and 3.4, $h(y,x)$ is identified up to a level shift, and $g(w)$ is identified. By (5), the counterfactuals $Y(x)[w]$ are identified.*

*Proof.* Fix some $y_0$ in the support of $Y$ under $P$ and fix some value $w$. We note that by Assumptions 3.1 and 3.2

$$g(w) - (h(Y[w],X) - h(y_0,X)) = h(y_0,X) + h(Y(x_0)[w_0],x_0)$$

The left-hand side is a known function $Q(w, X)$ by Assumption 3.3. By Assumption 3.4,

$$\mathbb{E}\left[h(Y(x_0)[w_0], x_0) \mid Z\right] = \mathbb{E}\left[h(Y(x_0)[w_0], x_0)\right] = C_{P^*}$$

is a constant. Thus, consider the integral equation

$$\mathbb{E}[v(X) \mid Z] = 0.$$

The choice $v(x) = h(y_0, x) - Q(w, x) + C_{P^*}$ solves the integral equation. Since it is the only solution by Assumption 3.4, the distribution $P$ identifies $h(y_0, x) - Q(w, x) + C_{P^*}$, which is equivalent to identifying $h(y_0, x)$ up to a level shift. By Assumption 3.3, $h(y, x) = (h(y, x) - h(y_0, x)) + h(y_0, x)$ is similarly identified up to a level shift. □

3.3. **Equivalence.** Berry and Haile (2024) impose the following core assumptions in their Assumptions 1–3. We present them in our notation with slight modifications reflecting their subsequent normalization.[11] They write

$$Y(x)[w] = \mathfrak{s}(w, x, \xi)$$

**Assumption 3.5** (Index). *$\mathfrak{s}(w, x, \xi) = \sigma(\gamma(w, \xi), x)$, where $\gamma$ is $J$-dimensional, and for all $j$, $\gamma_j(w, \xi) = g_j(w) + \xi_j$. For some fixed $w_0$, $g(w_0) = 0$ and $\frac{g(w_0)}{dw} = I_J$.*

**Assumption 3.6** (Invertible demand). *For all $x \in \mathcal{X}$, $\sigma(\cdot, x)$ is injective on the support of $\gamma(w, \xi)$.*

**Assumption 3.7** (Injective index). *For all $\xi$ in its support, $\gamma(\cdot, \xi)$ is injective on $\mathcal{W}$.*

Like Section 2, our main result in this section proves that the formulations are equivalent. The map $\sigma$ and the quantity $\xi$ are identified through the same argument Proposition 3.1 that identfies $Y(x)[w]$.

**Theorem 3.2.** *Assumptions 3.5 to 3.7 are equivalent to (the $J$-dimensional analogues of) Assumptions 3.1 and 3.2.*

*Proof.* Without essential loss of generality, we show the equivalence for $J = 1$. Suppose Assumptions 3.5 to 3.7 hold. Choose $h(y, x) = \sigma^{-1}(y, x)$, where $\sigma^{-1}$ is such that in Assumption 3.5

$$\sigma^{-1}(\mathfrak{s}(w, x, \xi), x) = \gamma(w, x).$$

The existence of $\sigma^{-1}$ is given by Assumption 3.6. Choose $g(w)$ as in $g(w)$ in Assumption 3.5, which is invertible by Assumption 3.7 and normalized appropriately. Now,

[11]Relative to Assumption 1 in Berry and Haile (2014), Assumption 3.5 normalizes the index directly, following their Section 2.5. Relative to their setting, we suppressed other market-level interventions (their $X_t$) that may enter $\gamma$, doing so makes the normalization in their Section 2.3 unnecessary, which we impose in Assumption 3.5 directly.

given these choices,

$$h(Y(x)[w_1], x_1) - h(Y(x)[w_0], x_0) = (g(w_1) + \xi) - (g(w_0) + \xi) = g(w_1) - g(w_0)$$

choice almost surely. This proves both Assumptions 3.1 and 3.2 by choosing $w_1, x_1, w_0, x_0$ appropriately.

For the reverse direction, let us impose Assumptions 3.1 and 3.2. By an analogue of (5), we can write

$$Y(x)[w] = h^{-1}\left(g(w) + h(Y(x_0)[w_0], x_0), x\right).$$

Set $\xi = h(Y(x_0)[w_0], x_0)$, $\gamma(w, \xi) = g(w) + \xi$, and $\sigma(\cdot, x) = h^{-1}(\cdot, x)$. These choices are appropriately invertible by assumption. This proves Assumptions 3.5 to 3.7. □

## 4. Conclusion

This paper connects fundamental results in demand estimation (Berry and Haile, 2014, 2024) to the Neyman–Rubin potential outcomes model. We find that the index and invertibility restrictions in Berry and Haile (2014, 2024) can be equivalently stated and interpreted as (i) certain causal effects may be purged by transforming the outcome and (ii) the transformed outcome satisfies either a form of homogeneous linear effects or a form of individual parallel trends. While strong—implying homogeneous treatment effects—these assumptions retain considerable flexibility because the transformations are latent. This reformulation is nontrivial in the sense that it makes no reference to latent variables beyond the potential outcomes themselves. We present analogues of identification arguments—largely following Berry and Haile (2014, 2024)—in terms of these equivalent assumptions. We also derive conditions under which Berry and Haile's (2014) demand model satisfies causally correct specification (Andrews *et al.*, forthcoming).

## References


Abdulkadiroğlu, A., Angrist, J. D., Narita, Y. and Pathak, P. (2022). Breaking ties: Regression discontinuity design meets market design. *Econometrica*, **90** (1), 117–151. 5

Andrews, I., Barahona, N., Gentzkow, M., Rambachan, A. and Shapiro, J. M. (forthcoming). Structural estimation under misspecification: Theory and implications for practice. *Quarterly Journal of Economics*. 3, 9, 10, 11, 17

Angrist, J. D., Graddy, K. and Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies*, **67** (3), 499–527. 2

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, **82** (5), 1749–1797. 1, 2, 3, 4, 7, 8, 9, 16, 17

— and — (2021). Foundations of demand estimation. In *Handbook of industrial organization*, vol. 4, Elsevier, pp. 1–62. 4, 7, 8, 13, 14

— and — (2024). Nonparametric identification of differentiated products demand using micro data. *Econometrica*, **92** (4), 1135–1162. 1, 2, 3, 12, 13, 14, 15, 16, 17

Blandhol, C., Bonney, J., Mogstad, M. and Torgovitsky, A. (2022). *When is TSLS actually late?* Tech. rep., National Bureau of Economic Research Cambridge, MA. 5

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, **91** (6), 2155–2185. 11

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, **16** (3), 199–231. 2

Card, D. (2022). Who set your wage? *American Economic Review*, **112** (4), 1075–1090. 3

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, **139** (2), 891–936. 5

Conlon, C. and Mortimer, J. H. (2021). Empirical properties of diversion ratios. *The RAND Journal of Economics*, **52** (4), 693–726. 3

Currie, J., Kleven, H. and Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. In *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 110, pp. 42–48. 3

Deaner, B. (forthcoming). The trade-off between flexibility and robustness in instrumental variables analysis. *American Economic Review*. 12

Egan, M. L., MacKay, A. and Yang, H. (2021). *What drives variation in investor portfolios? estimating the roles of beliefs and risk preferences*. Tech. rep., National Bureau of Economic Research. 3

Fisher, R. (1935). *The design of experiments*. Springer. 5

Goldsmith-Pinkham, P. (2024). *Tracking the Credibility Revolution across Fields*. Tech. rep. 3

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, NBER, pp. 475–492. 2

Humphries, J. E., Ouss, A., Stavreva, K., Stevenson, M. and van Dijk, W. (). Conviction, incarceration, and recidivism: Understanding the revolving door. 3

Imbens, G., Kallus, N., Mao, X. and Wang, Y. (2024). Long-term causal inference under persistent confounding via data combination. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, p. qkae095. 7

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** (2), 467–475. 2

Longuet-Marx, N. (2024). Party lines or voter preferences? explaining political realignment. In *2024 APPAM Fall Research Conference*, APPAM. 3

Mogstad, M. and Torgovitsky, A. (2024). Instrumental variables with unobserved heterogeneity in treatment effects. In *Handbook of Labor Economics*, vol. 5, Elsevier, pp. 1–114. 5

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, **71** (5), 1565–1578. 6, 15

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X. and Miao, W. (2024). An introduction to proximal causal inference. *Statistical Science*, **39** (3), 375–390. 7

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, **70** (1), 331–341. 2

## Appendix A. Miscellany

A.1. **Causally correct specification.** Here we provide an example in which there exists a nontrivial-in-$d$ $\tilde{h}(d)$ such that

$$\mathbb{E}[\tilde{h}(D) \mid Z, W] = \mathbb{E}[L_0(X_2) \mid Z, W]$$

for some choice of $L_0$. Suppose $Z > 0, W > 0$ and

$$(D, X_2) \mid (Z, W) \sim \mathcal{N}\left(\begin{bmatrix} ZW \\ W \end{bmatrix}, \Sigma\right)$$

is jointly Gaussian. First note that this choice does not sacrifice completeness since the conditional likelihood is linear exponential family, $(D, X_2)$ is complete sufficient for $\mu(Z, W) = (ZW, W)$, which ranges over an open set in $\mathbb{R}^2$.

Next, observe that by construction

$$\mathbb{E}[D - WX_2 \mid Z, W] = 0$$

Let $\tilde{h}(d) = d$ and let $L(w, x_2) = wx_2$. This choice satisfies the construction in Footnote 7