Online Appendix

Appendix A Proofs

Proof of Lemma 1. As we note in the text, in our parametric framework, the falsifiable restriction in Equation (28) of Berry and Haile (2014) is[35]

$$E[\omega_{mjt} \mid \mathrm{w}_{jt}, z_{jt}] \quad = \quad E[p_{jt} - \Delta_{mjt} - \bar{c}_{mj}(\mathrm{w}_{jt}) \mid \mathrm{w}_{jt}, z_{jt}] \quad = \quad 0 \qquad a.s.$$

Since observed prices are generated under the true model as

$$p_{jt} \quad = \quad \Delta_{0jt} + c_{0jt} \quad = \quad \Delta_{0jt} + \bar{c}_{0j}(\mathrm{w}_{jt}) + \omega_{0jt}$$

and $E[\omega_{0jt} \mid \mathrm{w}_{jt}, z_{jt}] = 0$ under Assumption 2, the falsifiable restriction is equivalent to

$$E[\Delta_{0jt} + \bar{c}_{0j}(\mathrm{w}_{jt}) + \omega_{0jt} - \Delta_{mjt} - \bar{c}_{mj}(\mathrm{w}_{jt}) \mid \mathrm{w}_{jt}, z_{jt}] \quad = \quad 0 \qquad a.s.$$

or equivalently

$$E[\Delta_{0jt} - \Delta_{mjt} \mid \mathrm{w}_{jt}, z_{jt}] \quad = \quad \bar{c}_{mj}(\mathrm{w}_{jt}) - \bar{c}_{0j}(\mathrm{w}_{jt}) \qquad a.s.$$

giving the result. □

Proof of Lemma 2. We prove the inverse of both directions. If the model is not falsified, then there exists a set of cost functions $\{\bar{c}_{mj}(\mathrm{w}_{jt})\}_j$ satisfying the falsifiable restriction. Since neither $\bar{c}_{mj}$ nor $\bar{c}_{0j}$ can depend on the instruments, this means (by Lemma 1) that for each j and each value of w_{jt}, the expectation $E[\Delta_{0jt} - \Delta_{mjt} | \mathrm{w}_{jt}, z_{jt}]$ is almost everywhere constant with respect to z_{jt}. Taking the limit

$$\lim_{h \to 0} \frac{E[\Delta_{0jt} - \Delta_{mjt} \mid \mathrm{w}_{jt}, z_{jt} = \tilde{z}_{jt} + h_k] - E[\Delta_{0jt} - \Delta_{mjt} \mid \mathrm{w}_{jt}, z_{jt} = \tilde{z}_{jt}]}{h}$$

as in the text and noting that this must be 0 almost surely, this becomes

$$E\left[\frac{\mathrm{d}\Delta_{0jt}}{\mathrm{d}z_{jt}^{(k)}} - \frac{\mathrm{d}\Delta_{mjt}}{\mathrm{d}z_{jt}^{(k)}} \mid \mathrm{w}_{jt}, z_{jt}\right] = 0 \qquad a.s.$$

giving the result.

[35] See Section 6, Case 2 in Berry and Haile (2014) for a discussion of their non-parametric environment.

For the opposite direction, if for every j and k, $E\left[\frac{\mathrm{d}\Delta_{0jt}}{\mathrm{d}z_{jt}^{(k)}} - \frac{\mathrm{d}\Delta_{mjt}}{\mathrm{d}z_{jt}^{(k)}} \mid \mathrm{w}_{jt}, z_{jt}\right] = 0$ almost surely, then $E[\Delta_{0jt} - \Delta_{mjt} \mid \mathrm{w}_{jt}, z_{jt}]$ must be the same for almost all values of z_{jt}. If so, define

$$\bar{c}_{mj}(\mathrm{w}_{jt}) \quad = \quad \bar{c}_{0j}(\mathrm{w}_{jt}) + E_{z_{jt}}\left[E[\Delta_{0jt} - \Delta_{mjt} \mid \mathrm{w}_{jt}, z_{jt}]\right]$$

and $\bar{c}_{mj}$ satisfies the equality condition in Lemma 1 almost surely, so the model is not falsified. □

Proof of Proposition 1. See text preceding Proposition 1.

□

Proof of Corollary 1. Let $z_{jt}^{(k)}$, the k-th instrument for product j, be the i^{th} cost shifter of rival product ℓ. Since our instruments are cost shifters, under Assumption 5, $\frac{\partial\Delta_{mjt}}{\partial z_{jt}^{(k)}}$ and $\frac{\partial\Delta_{0jt}}{\partial z_{jt}^{(k)}}$ are both 0. From Proposition 1, then, model m is falsified if for some (j, k),

$$E\left[\left(P_{mt}^{-1} - P_{0t}^{-1}\right)_j \frac{\mathrm{d}p_0}{\mathrm{d}z_{jt}^{(k)}} \mid \mathrm{w}_{jt}, z_{jt}\right] \neq 0 \qquad w.p.p.$$

Since the instrument $z_{jt}^{(k)}$ is a cost shifter of product $\ell \neq j$,

$$\frac{\mathrm{d}p_0}{\mathrm{d}z_{jt}^{(k)}} \quad = \quad \frac{\partial p_0}{\partial c_t}\frac{\partial c_t}{\partial z_{jt}^{(k)}} \quad = \quad P_{0t} e_\ell \frac{\partial \bar{c}_{0j}}{\partial \mathrm{w}_{jt}^{(i)}},$$

where e_ℓ is the ℓ-th vector of the canonical basis. As a result, model m is falsified if

$$E\left[\left(P_{mt}^{-1}P_{0t} - I\right)_j e_\ell \frac{\partial \bar{c}_{0j}}{\partial \mathrm{w}_{jt}^{(i)}} \mid \mathrm{w}_{jt}, z_{jt}\right] \neq 0 \qquad w.p.p.$$

for some j and some $\ell \neq j$. Since by assumption $\frac{\partial \bar{c}_{0j}}{\partial \mathrm{w}_{jt}^{(i)}} \neq 0$, if we choose ℓ and j such that the (j, ℓ) element of $E[P_{mt}^{-1}P_{0t} \mid \mathrm{w}_{jt}, z_{jt}]$ is nonzero, this condition holds and the model is falsified. □

Proof of Corollary 2. By Lemma 2, falsifiability comes down to whether for some j and k,

$$E\left[\frac{\mathrm{d}\Delta_{0jt}}{\mathrm{d}z_{jt}^{(k)}} - \frac{\mathrm{d}\Delta_{mjt}}{\mathrm{d}z_{jt}^{(k)}} \mid \mathrm{w}_{jt}, z_{jt}\right] \quad \neq \quad 0 \qquad w.p.p.$$

Let $z_{jt}^{(k)}$ be the i-th characteristic of product ℓ, and let $x_t^{(i)}$ denote the vector of that characteristic for all J products. Note that $x_t^{(i)}$ has both a direct effect on Δ_{mt} and an indirect effect through its impact on equilibrium prices,

$$\frac{\mathrm{d}\Delta_{mt}}{\mathrm{d}x_t^{(i)}} = \frac{\partial\Delta_{mt}}{\partial x_t^{(i)}} + \frac{\partial\Delta_{mt}}{\partial p_t}\frac{\mathrm{d}p_0}{\mathrm{d}x_t^{(i)}}$$

where $\frac{\mathrm{d}p_0}{\mathrm{d}x_t^{(i)}}$ is the effect of $x_t^{(i)}$ on equilibrium prices under the true model 0.

Under Assumption 8, $x_{jt}^{(i)}$ and p_{jt} affect Δ_{mt} and Δ_{0t} *directly* only through δ_{jt}, so

$$\frac{\partial\Delta_{mt}}{\partial x_t^{(i)}} = \frac{\partial\Delta_{mt}}{\partial\delta_t}\frac{\partial\delta_t}{\partial x_t^{(i)}} = \frac{\partial\Delta_{mt}}{\partial\delta_t}\beta^{(i)}I$$

and

$$\frac{\partial\Delta_{mt}}{\partial p_t} = \frac{\partial\Delta_{mt}}{\partial\delta_t}\frac{\partial\delta_t}{\partial p_t} = \frac{\partial\Delta_{mt}}{\partial\delta_t}(-\alpha I)$$

and, putting the two together,

$$\frac{\partial\Delta_{mt}}{\partial x_t^{(i)}} = -\frac{\beta^{(i)}}{\alpha}\frac{\partial\Delta_{mt}}{\partial p_t}$$

We already defined the notation $H_{\Delta_{mt}} = \frac{\partial\Delta_{mt}}{\partial p_t}$, so $\frac{\partial\Delta_{mt}}{\partial x_t^{(i)}} = -\frac{\beta^{(i)}}{\alpha}H_{\Delta_{mt}}$.

Next, to calculate $\frac{\mathrm{d}p_t}{\mathrm{d}x_t^{(i)}}$, recall that equilibrium prices are defined implicitly as the solution to the true first-order conditions $F(\cdot) = p_t - c_t - \Delta_{0t} = 0$. By the implicit function theorem,

$$\frac{\mathrm{d}p_t}{\mathrm{d}x_t^{(i)}} = -\left[\frac{\partial F}{\partial p_t}\right]^{-1}\left[\frac{\partial F}{\partial x_t^{(i)}}\right] = -\left[I - H_{\Delta_{0t}}\right]^{-1}\left[-\frac{\partial\Delta_{0t}}{\partial x_t^{(i)}}\right] = -\left[I - H_{\Delta_{0t}}\right]^{-1}\left[\frac{\beta^{(i)}}{\alpha}H_{\Delta_{0t}}\right]$$

Recalling that $P_{mt} = (I - H_{\Delta_{mt}})^{-1}$, this is

$$\frac{\mathrm{d}p_t}{\mathrm{d}x_t^{(i)}} = -\frac{\beta^{(i)}}{\alpha}P_{0t}\left(I - P_{0t}^{-1}\right) = \frac{\beta^{(i)}}{\alpha}(I - P_{0t})$$

Plugging these into $\frac{\mathrm{d}\Delta_{mt}}{\mathrm{d}x_t^{(i)}} = \frac{\partial\Delta_{mt}}{\partial x_t^{(i)}} + \frac{\partial\Delta_{mt}}{\partial p_t}\frac{\mathrm{d}p_t}{\mathrm{d}x_t^{(i)}}$ gives

$$\frac{\mathrm{d}\Delta_{mt}}{\mathrm{d}x_t^{(i)}} = -\frac{\beta^{(i)}}{\alpha}H_{\Delta_{mt}} + H_{\Delta_{mt}}\left(\frac{\beta^{(i)}}{\alpha}(I - P_{0t})\right) = -\frac{\beta^{(i)}}{\alpha}H_{\Delta_{mt}}P_{0t} = -\frac{\beta^{(i)}}{\alpha}(I - P_{mt}^{-1})P_{0t}$$

From this,

$$
\begin{aligned}
E\left[\frac{\mathrm{d}\Delta_{0jt}}{\mathrm{d}z_{jt}^{(k)}}-\frac{\mathrm{d}\Delta_{mjt}}{\mathrm{d}z_{jt}^{(k)}}\mid \mathrm{w}_{jt},z_{jt}\right] &= E\left[\frac{\beta^{(i)}}{\alpha}\left((P_{0t}^{-1}-P_{mt}^{-1})P_{0t}\right)_{j,\ell}\mid \mathrm{w}_{jt},z_{jt}\right] \\
&= \frac{\beta^{(i)}}{\alpha}E\left[\left(I-P_{mt}^{-1}P_{0t}\right)_{j,\ell}\mid \mathrm{w}_{jt},z_{jt}\right]
\end{aligned}
$$

Thus, unless $E[(P_{mt}^{-1}P_{0t})_j \mid \mathrm{w}_{jt},z_{jt}]=e_j'$ for each j almost surely, there is some (j,k) satisfying $E\left[\frac{\mathrm{d}\Delta_{0jt}}{\mathrm{d}z_{jt}^{(k)}}-\frac{\mathrm{d}\Delta_{mjt}}{\mathrm{d}z_{jt}^{(k)}}\mid \mathrm{w}_{jt},z_{jt}\right]\neq 0$ w.p.p., the condition for falsifiability under Lemma 2. □

Proof of Proposition 2. Given (3), $c_{0jt}=\bar{c}_{0j}(\mathrm{w}_{jt})+\omega_{0jt}=\nu_t p_{jt}-\tau_t-\nu_t\Delta_{0jt}$ and $c_{mjt}=\bar{c}_{mj}(\mathrm{w}_{jt})+\omega_{mjt}=\nu_t p_{jt}-\tau_t-\nu_t\Delta_{mjt}$, so

$$
\begin{aligned}
\omega_{mjt} &= \nu_t p_{jt}-\tau_t-\nu_t\Delta_{mjt}-\bar{c}_{mj}(\mathrm{w}_{jt})+\bar{c}_{0j}(\mathrm{w}_{jt})+\omega_{0jt}-\nu_t p_{jt}+\tau_t+\nu_t\Delta_{0jt} \\
&= \nu_t\Delta_{0jt}-\nu_t\Delta_{mjt}+\bar{c}_{0j}(\mathrm{w}_{jt})-\bar{c}_{mj}(\mathrm{w}_{jt})+\omega_{0jt}
\end{aligned}
$$

Since $\bar{c}_{0j}$, $\bar{c}_{mj}$, and $E(\omega_{0jt}|\mathrm{w}_{jt},z_{jt})$ don't depend on the instruments, falsification occurs if $E(\nu_t\Delta_{0jt}-\nu_t\Delta_{mjt}|\mathrm{w}_{jt},z_{jt})$ for some j varies with one of the instruments. Since

$$
\frac{\mathrm{d}(\nu_t\Delta_{0jt})}{\mathrm{d}z_{jt}^{(k)}} = \nu_t\frac{\partial\Delta_{0jt}}{\partial z_{jt}^{(k)}}+\nu_t\frac{\partial\Delta_{0jt}}{\partial p_t}\frac{\mathrm{d}p_0}{\mathrm{d}z_{jt}^{(k)}}+\Delta_{0jt}\frac{\mathrm{d}\nu_t}{\mathrm{d}z_{jt}^{(k)}}
$$

and $\frac{\partial\Delta_{0jt}}{\partial p_t}=\frac{\partial(p_t-c_{0t})}{\partial p_t}=I-P_{0t}^{-1}$, and the analogous expression for $\frac{\mathrm{d}(\nu_t\Delta_{mjt})}{\mathrm{d}z_{jt}^{(k)}}$, (4) follows. □

Proof of Corollary 3. Since the instrument is $z_t=\tau_t$, $\frac{\mathrm{d}\nu_t}{\mathrm{d}z_{jt}^{(k)}}=0$; and under Assumption 9, the tax rate doesn't enter into Δ_{mjt} or Δ_{0jt} directly, so falsification obtains if

$$
E\left[\nu_t(P_{mt}^{-1}-P_{0t}^{-1})_j\frac{\mathrm{d}p_0}{\mathrm{d}\tau_t}\middle|\, \mathrm{w}_{jt},\tau_t\right] \neq 0
$$

with positive probability for some j. The unit tax is the same as an increase in marginal costs for every product, so $\frac{\mathrm{d}p_0}{\mathrm{d}\tau_t}=\sum_j\frac{\mathrm{d}p_0}{\mathrm{d}c_{jt}}=P_0\iota$, where ι is a vector of ones. Falsification therefore requires that for some j, with positive probability over observables,

$$
\begin{gathered}
E\left[\nu_t(P_{mt}^{-1}-P_{0t}^{-1})_j P_{0t}\iota\middle|\, \mathrm{w}_{jt},\tau_t\right] \neq 0 \\
\updownarrow \\
E\left[\nu_t(P_{mt}^{-1}P_{0t}-I)_j\iota\middle|\, \mathrm{w}_{jt},\tau_t\right] \neq 0
\end{gathered}
$$

or the elements of the j^{th} row of $P_{mt}^{-1}P_{0t}$ don't sum to 1. □

Proof of Corollary 4. With the added assumption that either $P_{mt}^{-1} = P_{0t}^{-1}$ or $\frac{\mathrm{d}p_0}{\mathrm{d}z_t} = 0$, the first term in Equation 4 vanishes; so with $z_t = \nu_t$, under Proposition 2, falsification obtains if and only if for some j, $E\left[\Delta_{0jt} - \Delta_{mjt} \middle| \mathrm{w}_{jt}, \nu_t\right] \neq 0$ w.p.p., giving the result. □

Proof of Corollary 5. With ν_t fixed, government revenue at unit tax rate τ is $R_0(\tau) = \tau \sum_j s_{jt}(p_{0t}(\tau)) + (1-\nu_t)\sum_j p_{0t}(\tau)s_{jt}(p_{0t}(\tau))$; let $R_m(\tau)$ be the predicted revenue under model m, based on predicted prices $p_{mt}(\tau)$. As noted in the text, $\frac{dp_{mt}}{d\tau} = \frac{1}{\nu_t}P_{mt}\iota$ for either model. If $P_{mt}\iota = P_{0t}\iota$ everywhere and model m and the truth give the same prices at the observed tax rate $\tau = \tau_t$, then they predict the same prices $p_{mt}(\tau) = p_{0t}(\tau)$ at every tax rate, and therefore the same market shares $s_{jt}(p_{mt}(\tau)) = s_{jt}(p_{0t}(\tau))$ and revenue $R_m(\tau) = R_0(\tau)$. □

Proof of Corollary 6. With τ_t fixed, government revenue under ad valorem tax rate ν is $R_0(\nu) = \tau_t \sum_j s_{jt}(p_{0t}(\nu)) + (1-\nu)\sum_j p_{0t}(\nu)s_{jt}(p_{0t}(\nu))$; let $R_m(\nu)$ be the predicted revenue under model m. As noted in the text, for either model, $\frac{\mathrm{d}p_{mt}}{d\nu} = \frac{1}{\nu}P_{mt}(\Delta_{mt} - p_{mt})$. If $P_{mt}(\Delta_{mt} - p_{mt}) = P_{0t}(\Delta_{0t} - p_{0t})$ whenever $p_{mt} = p_{0t}$ and $p_{mt} = p_{0t}$ at the observed tax rate $\nu = \nu_t$, then $p_{mt}(\nu) = p_{0t}(\nu)$ for all tax rates, and the two models therefore predict the same market shares and revenue at every tax rate. □

Appendix B Connecting Falsification and Point Identification when Models are Nested

The results in Sections 3 and 4 are cast in terms of falsification of a candidate model m, and we pursue a testing approach in our simulations and application, but the results in the paper are also useful to inform identification (and thus estimation) exercises. In particular, our results can guide ex-ante instrument selection to avoid irrelevant instruments for estimation.

Similar to Magnolfi and Sullivan (2022), consider a setting where different candidate models belong to a parametric class, so that markups can be written as $\Delta(\theta)$ for a vector of parameters $\theta \in \Theta$. In this class of models, falsification proceeds as in our general discussion in Sections 3 and 4. In particular, a model m corresponding to a parameter value θ_m, is falsified by instruments z_{jt} under the general conditions laid out in Proposition 1. Within the parametric class of nested models, however, we can also discuss the identification of the true model, characterized by θ_0. For any value of θ, let the implied cost be $p_{jt} - \Delta_{jt}(\theta) = c(\mathrm{w}_{jt};\theta) = \bar{c}(\mathrm{w}_{jt};\theta) + \omega_{jt}(\theta)$. We say that the true model is *point identified by the instruments* z_{jt} when $E[\omega_{jt}(\theta)|\mathrm{w}_{jt}, z_{jt}] = 0$ if and only if $\theta = \theta_0$. Thus, point identification requires that all models m for which $\theta_m \neq \theta_0$ are falsified by the instruments z_{jt}. Conversely, when there exist a a model m such that $\theta_m \neq \theta_0$ which is not falsified by z_{jt}, θ_0 is not point identified by these instruments. Clearly, instruments that ensure falsification for all incorrect models in the class are prime candidates to be used for identification and estimation. However, instruments for which an incorrect model is not falsified would lead to a failure of point identification.

As an example of a conduct model defined by a continuous parameter, IO economists often use profit weights as a reduced-form way to model collusion or common ownership (e.g., Backus et al. (2021)). We illustrate how our framework for understanding falsification through pass-through can be useful for ex-ante instrument selection for identification and estimation with two examples based on such profit weights.

Example 8: We remain in the example environment of the paper – two single-product firms and logit demand. Suppose the two firms compete in quantities a la Cournot, but instead of maximizing its own profit, each firm maximizes a weighted sum of its own and the other firm's profits. Specifically, each firm j chooses quantity s_{jt} to solve

$$\max_{s_{jt}} \left\{ s_{jt}(p_{jt}(\cdot) - c_{jt}) + \theta_j s_{-jt}(p_{-jt}(\cdot) - c_{-jt}) \right\}$$

where $-j$ refers to the identity of the rival firm. This nests standard Cournot competition (when $\theta_1 = \theta_2 = 0$) and perfect collusion/joint profit maximization (when $\theta_1 = \theta_2 = 1$), along with intermediate cases that might be interpreted as "imperfect collusion". The corresponding markups under logit demand are

$$\Delta_{WCt}(\theta) = \begin{bmatrix} \frac{1-(1-\theta_1)s_{2t}}{\alpha s_{0t}} \\ \frac{1-(1-\theta_2)s_{1t}}{\alpha s_{0t}} \end{bmatrix},$$

with WC standing for the Weighted Cournot model.

For this setting, our falsification framework can be used to show that either cost or demand side instruments allow point identification of the profit weights $\theta = (\theta_1, \theta_2)$:

Result 1. *If the data are generated by Cournot competition with profit weight equal to* $\theta = \theta_0 \in [0,1]^2$*, then either cost side or demand side instruments point identify the true parameter.*

To see this result, we first simplify the first-order conditions and work out the pass-through and inverse pass-through matrices

$$P_{WCt}^{-1} = \frac{1}{s_{0t}} \begin{bmatrix} 1-s_{2t} & \theta_1 s_{2t} \\ \theta_2 s_{1t} & 1-s_{1t} \end{bmatrix} \quad \text{and} \quad P_{WCt} = \frac{s_{0t}}{\kappa_{WCt}} \begin{bmatrix} 1-s_{1t} & -\theta_1 s_{2t} \\ -\theta_2 s_{1t} & 1-s_{2t} \end{bmatrix}$$

where $\kappa_{WCt} = (1-s_{1t})(1-s_{2t}) - \theta_1\theta_2 s_{1t}s_{2t}$.

Suppose the true model is weighted Cournot competition with weights $\theta_0 = (\theta_{01}, \theta_{02})$, and we are interested in falsifying a model of weighted Cournot with misspecified weights $\theta_m = (\theta_{m1}, \theta_{m2}) \neq \theta_0$. Focusing on the off-diagonal terms and dropping constants, we can calculate

$$P_{WCmt}^{-1} P_{WC0t} \quad \propto \quad \begin{bmatrix} \star & (\theta_{m1}-\theta_{01})s_{2t}(1-s_{2t}) \\ (\theta_{m2}-\theta_{02})s_{1t}(1-s_{1t}) & \star \end{bmatrix}$$

Thus, if $\theta_{m1} > \theta_{01}$, the top-right off-diagonal is always positive, and therefore positive when its expectation is taken over unobservables; if $\theta_{m1} < \theta_{01}$, it's always negative, hence negative in expectation. Likewise, if $\theta_{m2} > \theta_{02}$, the bottom-right term is always positive, and if $\theta_{m2} < \theta_{02}$ always negative. Thus, if $\theta_m \neq \theta_0$, the matrix $[P_m^{-1}P_0]^*$ is not diagonal, so under Corollary 1, any incorrect model within the class is falsified by cost side instruments under Corollary 1. As for demand side instruments, the profit-weighted Cournot model satisfies Assumption 7 and logit demand satisfies Assumption 8, so since $[P_m^{-1}P_0]^*$ is not the identity matrix, demand side instruments falsify any incorrect model under Corollary 2. As a result, either type of instrument allows for point identification of θ_0. •

Example 9: In the same environment, suppose now that the two firms compete in prices with profit weights. For simplicity, suppose the two firms' profit weights are the same, so each firm j chooses price p_{jt} to solve

$$\max_{p_{jt}} \left\{ (p_{jt}-c_{jt})s_{jt}(p_t) + \theta(p_{-jt}-c_{-jt})s_{-jt}(p_t) \right\}$$

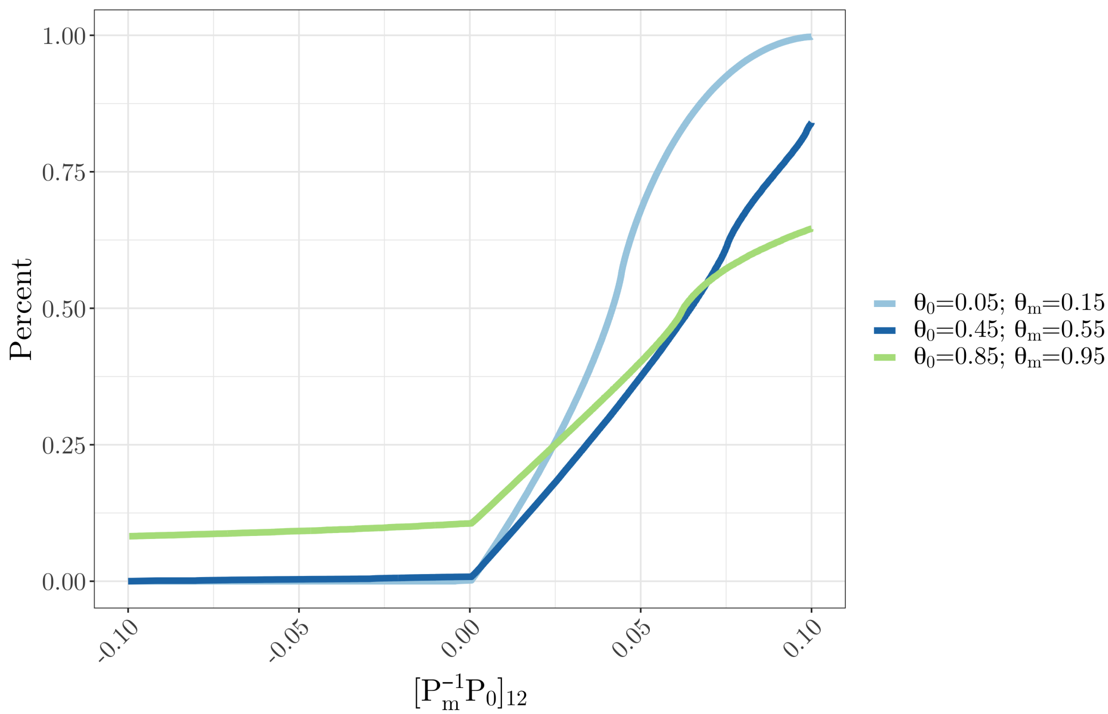
Under logit demand, the two firms' first-order conditions give the markup function

$$\Delta_{WBt}(\theta) \quad = \quad \begin{bmatrix} \frac{1-(1-\theta)s_{2t}}{\alpha s_{0t}+\alpha(1-\theta^2)s_{1t}s_{2t}} \\ \\ \frac{1-(1-\theta)s_{1t}}{\alpha s_{0t}+\alpha(1-\theta^2)s_{1t}s_{2t}} \end{bmatrix}$$

with WB standing for the Weighted Bertrand model.

While we obtain identification in Weighted Cournot, irrespective of the value of θ_0 and of other observables, in the case of Weighted Bertrand, results are more nuanced. For a given

FIGURE 6: CDF of $[P_{mt}^{-1}P_{0t}]_{12}$ for Betrand Profit Weight Models



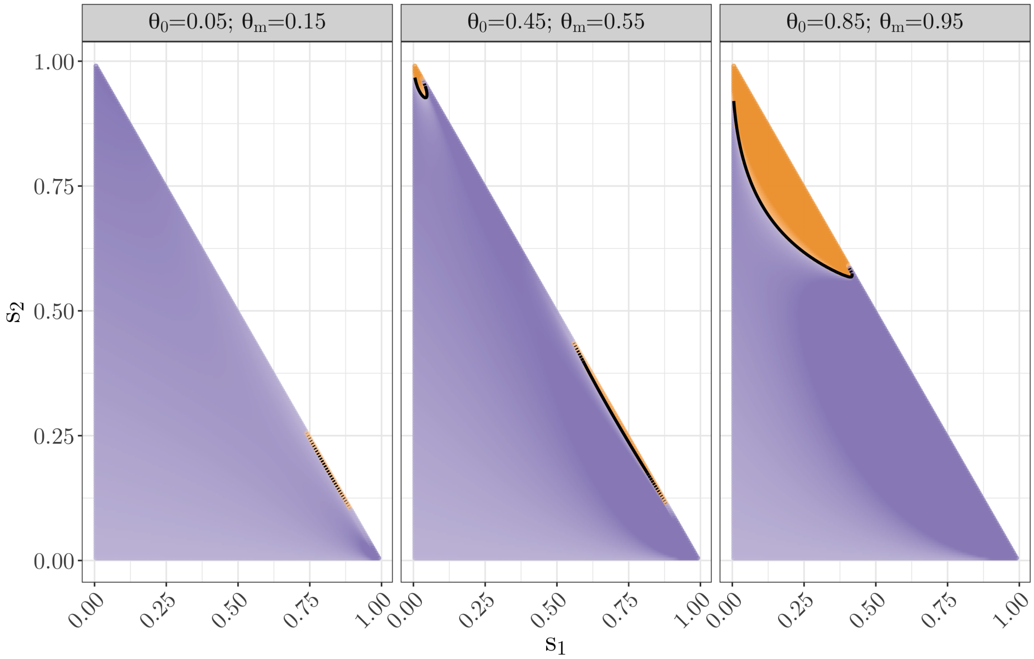
This figure plots distributions of $[P_{mt}^{-1}P_{0t}]_{12}$ over realizations of (s_{1t}, s_{2t}) for three different values of (θ_0, θ_m).

true value of θ_0 and a given misspecified model $\theta_m \neq \theta_0$, the off-diagonal terms of the matrix $P_m^{-1}P_0$ can be either positive or negative, but are not typically zero; and the diagonal terms can be greater than or less than 1, but are not typically equal to either 1 or to each other. This means there are realizations of observables (w_t, z_t) for which $[P_m^{-1}P_0]^\star$ could in principle be diagonal, but it won't typically be – for example, because some off-diagonal element of $P_{mt}^{-1}P_{0t}$ varies continuously and could take either sign, so its expectation being exactly zero requires a non-generic distribution of unobservables. Still, falsification could in principle fail, and therefore identification could fail as well.

To illustrate this possibility, we compute the top-right off-diagonal term, $[P_m^{-1}P_0]_{12}$, for a large sample of uniform draws of (s_{1t}, s_{2t}), and display the resulting distribution in Figure 6 for three representative pairs of true and misspecified models (θ_0, θ_m). While the distributions all have support that includes 0, the distributions are continuous, so getting a conditional expectation $E([P_{mt}^{-1}P_{0t}]_{12}) = 0$ would only hold for a non-generic set of observables.

We can also examine the value of the off-diagonal term at each point in the simplex of realizations of market shares (s_{1t}, s_{2t}). In each panel of Figure 7, we do so for three pairs of θ_0 and θ_m such that $\theta_m - \theta_0 = 0.1$. For each realization on the simplex, we indicate the value of $[P_m^{-1}P_0]_{12}$ with color gradients – darker orange indicates negative values that are larger in magnitude and darker purple indicates positive values that are larger in magnitude.

FIGURE 7: Values of $[P_{mt}^{-1}P_{0t}]_{12}$ for Realizations of Market Shares in Bertrand Profit Weight Models



This figure plots magnitudes of $[P_{mt}^{-1}P_{0t}]_{12}$ over the simplex of (s_{1t}, s_{2t}) for three different values of (θ_0, θ_m). Darker purple (orange) shading indicates more positive (negative) values. Black indicates a zero value.

The points where $[P_m^{-1}P_0]_{12} = 0$ are indicated in black. Immediately, one sees that for the vast majority of realizations of shares, $[P_m^{-1}P_0]_{12} > 0$. In fact, $[P_m^{-1}P_0]_{12}$ is always positive whenever the market share of both products is below 0.6. For an empirically relevant example, take the setting in Miller and Weinberg (2017). There, the outside option is defined in such a way that the market share of any product is less than 0.5 in all markets. In that case, the average of $[P_m^{-1}P_0]_{12}$ is positive in all three panels, and falsification of the wrong profit weight is possible with cost or demand side instruments.

As in the previous example, the conduct models satisfy Assumption 7 and the demand system satisfies Assumption 8, so $[P_m^{-1}P_0]^*$ being non-diagonal suffices for falsification of the wrong model. Thus, for either cost or demand side instruments, point identification is not theoretically guaranteed for a particular realization of observables; but with variation in observables, seems virtually guaranteed, as the knife-edge result of positive and negative values of $[P_m^{-1}P_0]_{12}$ cancelling out in expectation seems impossible across different realizations of observables. •

Appendix C Markup Assumption

Assumption 7 holds naturally for a wide range of models where firms choose actions to maximize profits. We suppress the market index t and suppose for simplicity that products $i = 1$ through f are sold by the same firm, and that the firm chooses a set of actions $\{a_i\}_{i=1}^{f}$ to maximize profits,

$$\max_{\{a_i\}, i \in f} \sum_{i \in f} (p_i(a) - c_i) s_i(a)$$

where prices $p(\cdot)$ and market shares $s(\cdot)$ are determined by the actions taken by all firms. First-order conditions are then

$$\sum_{i=1}^{f} \frac{\partial p_i}{\partial a_j} s_i + \sum_{i=1}^{f} (p_i - c_i) \frac{\partial s_i}{\partial a_j} \quad = \quad 0$$

or

$$\begin{bmatrix} \frac{\partial s_1}{\partial a_1} & \frac{\partial s_2}{\partial a_1} & \cdots & \frac{\partial s_f}{\partial a_1} \\ \frac{\partial s_1}{\partial a_2} & \frac{\partial s_2}{\partial a_2} & \cdots & \frac{\partial s_f}{\partial a_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_1}{\partial a_f} & \frac{\partial s_2}{\partial a_f} & \cdots & \frac{\partial s_f}{\partial a_f} \end{bmatrix} \begin{bmatrix} p_1 - c_1 \\ p_2 - c_2 \\ \vdots \\ p_f - c_f \end{bmatrix} = - \begin{bmatrix} \sum_{i \in f} s_i \frac{\partial p_i}{\partial a_1} \\ \sum_{i \in f} s_i \frac{\partial p_i}{\partial a_2} \\ \vdots \\ \sum_{i \in f} s_i \frac{\partial p_i}{\partial a_f} \end{bmatrix}$$

Stacking across firms, we then get

$$\left[\Omega \odot \left[\frac{\partial s}{\partial a} \right]' \right] \Delta \quad = \quad - \left[\Omega \odot \left[\frac{\partial p}{\partial a} \right]' \right] s$$

where Ω is the ownership matrix,[36] and therefore

$$\Delta \quad = \quad - \left[\Omega \odot \left[\frac{\partial s}{\partial a} \right]' \right]^{-1} \left[\Omega \odot \left[\frac{\partial p}{\partial a} \right]' \right] s$$

Note that the right-hand side has no room for costs or product characteristics to enter directly – it's all just ownership structure and the way that firm actions a map to market outcomes (p, s), which depends on the demand system.

(Within this more general model, Bertrand is just the special case where firms choose

[36] This is defined as $\Omega_{ij} = 1$ if products i and j are sold by the same firm, and zero otherwise.

prices, so $p(a) = a$ and therefore $\frac{\partial p}{\partial a} = I$; and Cournot is the special case where firms choose quantities so $s(a) = a$ and $\frac{\partial s}{\partial a} = I$. Here we're being more general about what exactly firms are choosing, and therefore what exactly they're assuming other firms are holding fixed while they optimize.)

This assumption also holds if firms maximize any weighted sum of their own profits, other firms' profits, and consumer surplus (or total welfare). Suppose the firm selling product j maximizes

$$\sum_{i=1}^{J} \gamma_{ji}(p_i - c_i)s_i + \lambda_j CS$$

where γ_{ji} is the weight the firm puts on the profits from product i (whether or not i is one of the same firm's products) and CS is consumer surplus. The first-order condition with respect to action a_j is then

$$\sum_{i=1}^{J} \gamma_{ji} \frac{\partial p_i}{\partial a_j} s_i + \sum_{i=1}^{J} \gamma_{ji}(p_i - c_i) \frac{\partial s_i}{\partial a_j} - \lambda_j s_j \quad = \quad 0$$

or, stacking and rearranging,

$$\Delta \quad = \quad \left[\Gamma \odot \left[\frac{\partial s}{\partial a} \right]' \right]^{-1} \left[\Lambda - \Gamma \odot \left[\frac{\partial p}{\partial a} \right]' \right] s$$

where Γ is a matrix of the γ_{ji} terms and Λ is a diagonal matrix of the λ_j terms. Once again, the right-hand side contains only constants and features of the demand system, not costs or product characteristics.

Finally, consider a market with some first-movers and some second-movers. To avoid getting bogged down in notation, we show the result for two single-product firms facing general demand, but the intuition is the same more generally. Conditional on the action a_1 chosen by the first-mover, the second-mover chooses a_2 to maximize $(p_2 - c_2)s_2$, giving first-order condition

$$(p_2 - c_2) \frac{\partial s_2}{\partial a_2} + \frac{\partial p_2}{\partial a_2} s_2 \quad = \quad 0$$

Defining $F(a_1, a_2)$ as the left-hand side, then, a_2 is implicitly defined as a function of a_1 as the solution to $F(a_1, a_2) = 0$, so by the implicit function theorem,

$$a_2'(a_1) \quad = \quad -\frac{\frac{\partial F}{\partial a_1}}{\frac{\partial F}{\partial a_2}} \quad = \quad -\frac{\frac{\partial p_2}{\partial a_1}\frac{\partial s_2}{\partial a_2} + (p_2 - c_2)\frac{\partial^2 s_2}{\partial a_2 \partial a_1} + \frac{\partial^2 p_2}{\partial a_1 \partial a_2} s_2 + \frac{\partial p_2}{\partial a_2}\frac{\partial s_2}{\partial a_1}}{\frac{\partial p_2}{\partial a_2}\frac{\partial s_2}{\partial a_2} + (p_2 - c_2)\frac{\partial^2 s_2}{\partial a_2^2} + \frac{\partial^2 p_2}{\partial a_2^2} s_2 + \frac{\partial p_2}{\partial a_2}\frac{\partial s_2}{\partial a_2}}$$

We can go a step further, rewriting firm 2's first-order condition as $p_2 - c_2 = -\frac{\partial p_2}{\partial a_2} s_2 \Big/ \frac{\partial s_2}{\partial a_2}$ and plugging that into the expression for a_2', to emphasize that a_2 depends only on features of the demand system (how p and s respond to a) and therefore not directly on costs. The first-mover's problem is

$$\max(p_1(a_1, a_2(a_1)) - c_1)s_1(a_1, a_2(a_1))$$

with first-order condition

$$\frac{\partial p_1}{\partial a_1} s_1 + \frac{\partial p_1}{\partial a_2} a_2' s_1 + (p_1 - c_1)\frac{\partial s_1}{\partial a_1} + (p_1 - c_1)\frac{\partial s_1}{\partial a_2} a_2' \quad = \quad 0$$

whence

$$p_1 - c_1 \quad = \quad -\frac{\frac{\partial p_1}{\partial a_1} s_1 + \frac{\partial p_1}{\partial a_2} a_2' s_1}{\frac{\partial s_1}{\partial a_1} + \frac{\partial s_1}{\partial a_2} a_2'}$$

We therefore have both firms' markups $p_j - c_j$ as functions of the demand system, with no place for marginal costs or product characteristics to enter directly. If we had infinite patience, we could make this same argument for the general model of many multi-product firms with some first- and some second-movers, and by induction, with more than two "rounds" of actions.

Appendix D Simulation Details

We provide further details on the simulation environment used in Section 6. Using the simulation class in PyBLP (Conlon and Gortmaker (2020)), we simulate data for 50,000 markets. In each market t, the number of products J_t is a randomly chosen integer between two to ten, leaving us with 319,719 observations in the sample. Each product j is produced by a single product firm.

For demand, we adopt a simple logit framework, in line with our falsification examples. Consumer i gets indirect utility of consuming product j in market t given by:

$$u_{ijt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where x_{jt} is a vector containing a constant and two observed product characteristics (x_{1jt} and x_{2jt}), p_{jt} is the price of the product, and ξ_{jt} and ϵ_{ijt} are unobservable shocks at the product-market and the individual product market level, respectively. The utility of the outside option is normalized to $u_{i0t} = \epsilon_{i0t}$. We draw each observed product characteristic x_{1jt} and x_{2jt} independently from the uniform distribution $U(0, 3)$, while ϵ_{ijt} is assumed to be distributed Type I extreme value. The distribution of the unobserved demand shocks ξ_{jt} is discussed below. The mean taste parameters β are set as $\beta = [1, 2, 1]$ while the price parameter $\alpha = -0.5$.

On the supply side, we assume that the marginal cost of producing product j in market t is $c_{jt} = \mathrm{w}_{jt}\gamma + \omega_{0jt}$ where where w_{jt} is a vector containing a constant and two observed cost shifters (w_{1jt} and w_{2jt}) which are excluded from demand. Marginal cost also depends on ω_{0jt} the true unobserved cost shock. As with the product characteristics, we draw each observed cost shifter w_{1jt} and w_{2jt} independently from the uniform distribution $U(0, 3)$. We adopt the default in PyBLP by drawing the unobserved demand and cost shocks ξ_{jt} and ω_{0jt} from a mean zero bivariate normal distribution with variances of 1 and a correlation of 0.9. We set $\gamma = [3, 0.5, 1.5]$. For simplicity, the market size is normalized to one for all t. The government levies both a unit tax (τ_t) and an ad valorem tax (v_t) on all products in market t. The unit tax is remitted by the firms while consumers remit the ad valorem tax. We draw the unit tax in each market from the uniform distribution $U(1, 2)$ while the ad valorem tax in each market is drawn from the uniform distribution $U(0, 0.2)$. We assume that the true model of conduct is Keystone pricing, whereby firms set tax exclusive prices as twice their marginal cost, or $\nu_t p_{jt} = 2c_{jt} + 2\tau_t$.

Appendix E Summary of IV Relevance in Examples

To help the reader, we now summarize the takeaways from the examples in the paper. Recall that we consider a stylized environment with two single-product firms, logit demand, and no unobservable variation in demand and cost. In the following table, for select combinations of true and alternative model, we indicate which of the sources of variation considered in the examples will permit falsification (✓), and which ones will not (×).

True	Tested	Cost Side	Demand Side	Unit Tax	Ad Valorem Tax
Bertrand	Keystone	✓	✓	✓	✓
	Cournot	✓	✓	✓	✓
	MC Pricing	✓	✓	✓	✓
Cournot	Keystone	×	✓	✓	✓
	MC Pricing	×	✓	✓	✓
Keystone	MC Pricing	×	×	✓	×
$\Delta_{0t} = \zeta_{0t}$	$\Delta_{mt} = \zeta_{mt}$	×	×	×	✓

Appendix F Additional Details: Data and Demand

F.1 Data Cleaning and Description

Table 4 summarizes the main data cleaning steps for our transaction-level data. We then take the cleaned transaction-level data and aggregate to the product-market level.

Table 4: Data Sample Cleaning Steps

Step	Sample Restriction	Sample Size	Resulting Revenue ($)
0	Original	74,427,564	1,213,615,962
1	Keep usable products	53,883,838	813,353,991
2	Drop package size $\neq$ inventory usable weight	53,791,104	809,548,311
3	Keep 1 and 3.5 g package sizes	34,418,816	490,387,068
4	Drop first 15 days of establishment's sales	34,286,246	488,555,343
5	Keep $\frac{\text{retail prices}}{\text{wholesale prices}} \in [1, 5]$	34,170,796	486,487,909
6	Keep weight $\leq$ 10 g	34,139,758	482,017,967

Table reports all steps to clean transaction-level data and their effect on sample size.

Table 5 provides descriptive statistics for the main variables in our database.

Table 5: Summary Statistics

	Mean	SD	Min	P25	Median	P75	Max
Price (per gram, tax inclusive)	11.04	3.05	4.54	9.00	10.89	12.86	21.07
Price (per gram, tax exclusive)	7.61	2.18	3.11	6.18	7.48	8.85	15.60
Shares (percent)	0.38	0.90	0.00	0.02	0.08	0.34	39.06
Wholesale Price (per gram)	3.60	0.91	1.50	3.00	3.58	4.09	6.50
Size	2.19	1.25	1.00	1.00	1.00	3.50	3.50
THC	5.37	7.90	0.00	0.53	0.99	6.80	62.66
CBD	0.42	1.56	0.00	0.00	0.09	0.22	36.93
Rival Products	443.67	571.17	1.00	60.00	196.00	499.00	2109.00
Firms by Market	13.80	15.55	2.00	3.00	7.00	16.00	51.00
Processors by Market	111.09	76.46	1.00	49.00	91.00	149.00	316.00
Local Tax Rate	0.089	0.007	0.070	0.085	0.088	0.095	0.103
Observations	153,936						

This table reports summary statistics for the main variables in our database.

F.2 Demand Estimation

We provide here further details on the demand system introduced in Section 7.2. In our demand system, each consumer i receives utility from product j in market t according to the indirect utility:

$$u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + F_{r(j)} + F_{\ell(j)} + F_{m(t)} + \xi_{jt} + \zeta_{it} + (1-\rho)\epsilon_{ijt}$$

where x_j includes a constant, package size, THC and CBD (and their values squared), and the log of the number of products offered in the store; we include this variable to capture variation in shelf space across stores. The variable p_{jt} is the price of product j in market t, and $F_{r(j)}$, $F_{\ell(j)}$, $F_{m(t)}$ denote fixed effects for the retailer selling product j, the processor producing product j, and the year-month of the retail transaction respectively. Consumer preferences for characteristics ($\beta_i = \bar{\beta} + \tilde{\beta} \times \text{Income}_i$) and price ($\alpha_i = \bar{\alpha} + \tilde{\alpha} \times \text{Income}_i$) vary with individual level income. ξ_{jt} and $\zeta_{it} + (1-\rho)\epsilon_{ijt}$ are unobservable shocks at the product-market and the individual product market level, respectively. Following the nested logit structure for our choice of nesting all inside goods together, ϵ_{ijt} is distributed Type 1 Extreme Value, and ζ_{it} is distributed according to the conjugate distribution (Cardell (1997)). To close the model we normalize consumer i's utility from the outside option as $u_{i0t} = \epsilon_{i0t}$. Given this utility specification, market shares s_{jt} as a function of observables, unobservables and parameters take on the standard form (Berry, Levinsohn, and Pakes (1995); Grigolon and Verboven (2014)).

Identification and Estimation: The identifying assumption for our demand model is that, for a vector of demand instruments z^d_{jt}, the moment condition $E[\xi_{jt} z^d_{jt}] = 0$. We construct several demand instruments. We first construct the number of products sold at competing dispensaries in market t to help identify ρ. Following Gandhi and Houde (2023) we also interact this instrument with the mean income in the market to help identify income interaction parameters. We further construct BLP-style instruments to capture the closeness of products in the product space. Specifically, in a market, we sum the amount of THC and CBD both for products within a given store and also for products in all other stores. We also include exogenous own cost shifters including the amount of rainfall and temperature in the region of production and their lags.

Because we do not take a stance on conduct, we perform demand estimation without using any supply-side moments.[37] However, we specify in Section 7.3 a menu of candidate models of conduct which includes Keystone pricing, Bertrand, and marginal cost pricing.

[37] See Appendix I of Duarte et al. (2024) for a comparison of sequential and simultaneous approach to conduct testing.

The set of instruments specified above includes shifters of prices (e.g., own cost shifters) that will be relevant under any conduct model in our menu including Keystone and marginal cost pricing.

TABLE 6: Demand Estimates

	(1) Logit-OLS		(2) Logit-2SLS		(3) RCNL	
	coef	s.e.	coef	s.e	coef	s.e
Price (in $)	-0.102	(0.002)	-0.283	(0.042)	-0.455	(0.066)
Package Size (= 3.5 oz)	0.441	(0.007)	0.335	(0.025)	0.110	(0.048)
THC	0.077	(0.003)	0.062	(0.005)	0.027	(0.007)
THC Squared	-0.003	(0.000)	-0.003	(0.000)	-0.001	(0.000)
CBD	-0.095	(0.009)	-0.084	(0.009)	-0.046	(0.010)
CBD Squared	-0.002	(0.001)	-0.002	(0.001)	-0.002	(0.000)
Log Number Own Products	-0.021	(0.016)	0.001	(0.017)	0.115	(0.027)
Constant	-5.791	(0.061)				
ρ					0.282	(0.059)
Income $\times$ Constant					-0.005	(0.003)
Income $\times$ Price					0.004	(0.002)
Median Own Price Elasticity	-1.113		-3.079		-6.451	
Median Aggregate Price Elasticity	-0.796		-2.192		-3.324	
Diversion to outside option	0.603		0.603		0.434	
Retailer FE	Yes		Yes		Yes	
Processor FE	Yes		Yes		Yes	
Year-Month FE	Yes		Yes		Yes	

Demand estimates for a logit model of demand obtained from OLS estimation are reported in column 1 and 2SLS estimation are reported in column 2. Column 3 reports estimates for the full RCNL demand model. Income is measured in $100,000. $n = 153,936$.

Results: Results for demand estimation are reported in Table 6. For reference, we also report, in Columns 1 and 2, estimates from a simple logit model. In Column 1, we estimate the model with ordinary least squares. In Column 2, we estimate the model via 2SLS, using the same instruments as in the main specification. The reduction in the price coefficient between Columns 1 and 2, indicates the presence of endogeneity not controlled for by the fixed effects. Column 3 reports estimates of the full demand model. Compared to columns

1 and 2, the full model yields a more elastic demand system with diversion to the outside option that departs from logit.

Appendix G Robustness Checks

We present here additional robustness results for our empirical application.

G.1 Robustness to Box-Cox Transformations of Income

Following Miravete et al. (2024) we incorporate in our demand system a more flexible specification of income effects by adopting a Box-Cox transformation. Specifically, we allow consumer i's price sensitivity parameter α_i to depend on a nonlinear transformation of income (y_i), so that

$$\alpha_i = \bar{\alpha} + \tilde{\alpha} \times y_i^{\lambda - 1}. \tag{6}$$

Our main specification corresponds to $\lambda = 2$. In principle, λ could be estimated as an additional parameter of the demand system. Because we lack the variation to credibly identify λ in our empirical environment, here we consider the robustness of our testing results to re-estimating our demand system for calibrated values of $\lambda \in \{0, 0.5, 1.5\}$. Table 7 reports demand estimates (in Panel A), and corresponding test results (in Panel B) for the three values of λ. While the nonlinear transformation of income makes a difference for some features of the demand system, test results are consistent with our main specification in Section 7, and strongly support Keystone as the better fitting model.

Table 7: Robustness to Box-Cox Transformations of Income in Demand

	RCNL Box Cox		
	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1.5$
Panel A: Demand Results:			
Price (in $)	-0.352	-0.377	-0.478
	(0.034)	(0.019)	(0.088)
Package Size (= 3.5 oz)	0.121	0.134	0.119
	(0.037)	(0.000)	(0.047)
THC	0.027	0.029	0.029
	(0.006)	(0.007)	(0.008)
THC Squared	-0.001	-0.001	-0.002
	(0.000)	(0.000)	(0.003)
CBD	-0.044	-0.046	-0.048
	(0.009)	(0.009)	(0.010)
CBD Squared	-0.001	-0.001	-0.002
	(0.001)	(0.001)	(0.006)
Log Number Own Products	0.141	0.135	0.110
	(0.024)	(0.025)	(0.027)
σ	0.391	0.376	0.274
	(0.057)	(0.000)	(0.060)
Income × Constant	0.005	0.006	-0.004
	(0.001)	(0.000)	(0.004)
Income × Price	0.421	0.421	0.008
	(0.395)	(0.019)	(0.005)
Median Own Price Elasticity	-5.612	-5.352	-6.238
Median Aggregate Price Elasticity	-2.394	-2.356	-3.258
Diversion to outside option	0.391	0.401	0.468
Retailer FE	Yes	Yes	Yes
Processor FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Panel B: Testing Results: Bertrand vs. Keystone, Tax Instruments			
T^{RV}	10.121	10.071	13.421
	***	***	***
F	38.5	32.8	43.0
	† † † ∧∧∧	† † † ∧∧∧	† † † ∧∧∧

Panel A reports demand estimates for a Box Cox transformation of income (measured in $100,000). Columns 1-4 correspond to different calibrated values of the λ parameter in Equation (6). Panel B reports, for each value of λ, the RV test statistics T^{RV} and the effective F-statistic (Duarte et al., 2024) for testing Bertrand versus Keystone with ad valorem tax instruments. A positive RV test statistic suggests a better fit of Keystone. The symbol $***$ indicates rejection of the null of equal fit 0.01 confidence level. The symbols † † † and $\wedge\wedge\wedge$ indicated that F is above the appropriate critical values for worst-case size below 0.075, and best-case power above 0.95, respectively. Both T^{RV} and the F-statistics account for two-step estimation error and clustering at the market level. $n = 153,936$.

G.2 Robustness of Test Results to Alternative Cost Specifications

We consider the robustness of our testing results in Table 3 to alternative specifications of marginal cost. Results are reported in Table 8. Panel A reproduces the specification in the main text. Panels B and C include package size indicators, and consider alternative specifications of fixed effects in marginal cost. For all specifications, our preferred tax instruments are strong for size and power, and conclude for superior fit of the Keystone model. Different specifications of marginal cost affect the variation that is available for testing conduct. In particular, the specifications in Panels B and C incorporate both geographic and time fixed effects, thus absorbing considerable variation. This weakens instruments, especially RC and PC. Despite these additional hurdles, whenever the null is rejected, the test always rejects in favor of Keystone.

TABLE 8: Test Results for Different Levels of Fixed Effects

	Instruments:			
Statistic	Tax	RC	PC	WP
Panel A: Product Fixed Effects				
T^{RV}	13.39	4.52	9.88	16.79
	$***$	$***$	$***$	$***$
F	40.6	2.2	15.6	1,591.6
	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$
Panel B: Retailer, Processor, and Year-Month Fixed Effects				
T^{RV}	2.39	-0.90	2.24	9.98
	$***$		$***$	$***$
F	4.2	0.5	1.4	4,912.6
	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$	$\dagger\dagger\dagger$	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$
Panel C: Retailer, and Year-Month Fixed Effects				
T^{RV}	2.74	-0.77	2.45	27.35
	$***$		$***$	$***$
F	4.1	0.5	1.4	3,232.1
	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$	$\dagger\dagger\dagger$	$\dagger\dagger\dagger$ $\wedge\wedge$	$\dagger\dagger\dagger$ $\wedge\wedge\wedge$

The table reports, for each set of instruments, the RV test statistics T^{RV} and the effective F-statistic (Duarte et al., 2024) for testing Bertrand versus Keystone. Different panels correspond to different levels of fixed effects in marginal cost. A positive RV test statistic suggests a better fit of Keystone. The symbol $***$ indicates rejection of the null of equal fit 0.01 confidence level. The symbols $\dagger\dagger\dagger$ and $\wedge\wedge\wedge$ indicated that F is below the appropriate critical values for worst-case size below 0.075, and best-case power above 0.95, respectively. Both T^{RV} and the F-statistics account for two-step estimation error and clustering at the market level. $n = 153,936$.

Appendix References

Backus, M., C. Conlon, and M. Sinkinson (2021): "Common Ownership and Competition in the Ready-To-Eat Cereal Industry," NBER Working Paper #28350.

Berry, S. and P. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Cardell, N. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13, 185–213.

Conlon, C. and J. Gortmaker (2020): "Best Practices for Differentiated Products Demand Estimation with PyBLP," *RAND Journal of Economics*, 51, 1108–1161.

Duarte, M., L. Magnolfi, M. Sølvsten, and C. Sullivan (2024): "Testing Firm Conduct," *Quantitative Economics*, 15, 571–606.

Gandhi, A. and J.-F. Houde (2023): "Measuring Substitution Patterns in Differentiated Products Industries," Working Paper.

Grigolon, L. and F. Verboven (2014): "Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation," *Review of Economics and Statistics*, 96, 916–935.

Magnolfi, L. and C. Sullivan (2022): "A Comparison of Testing and Estimation of Firm Conduct," *Economics Letters*, 212, 110316.

Miller, N. and M. Weinberg (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85, 1763–1791.