Estimation of BLP models with high-dimensional controls

Hua Jin

UCL

November 3, 2026

Motivation

BLP (Berry, Levinsohn, and Pakes, 1995) is widely used in IO to map consumer and product characteristics to demand

- Standard framework: a fixed, low-dimensional set of characteristics

Motivation

BLP (Berry, Levinsohn, and Pakes, 1995) is widely used in IO to map consumer and product characteristics to demand

- Standard framework: a fixed, low-dimensional set of characteristics

Modern data: many candidate characteristics, possibly more than markets [Compiani et al., 2026, Gillen et al., 2019]

Motivation

BLP (Berry, Levinsohn, and Pakes, 1995) is widely used in IO to map consumer and product characteristics to demand

- Standard framework: a fixed, low-dimensional set of characteristics

Modern data: many candidate characteristics, possibly more than markets [Compiani et al., 2026, Gillen et al., 2019] **Trade-off:**

- Omit controls $\Rightarrow$ omitted-variable bias
- Include many $\Rightarrow$ GMM imprecise or infeasible
- Select, then plug in $\Rightarrow$ invalidate inference [Chernozhukov et al., 2018]

Why high-dimensional controls?

Paper	**Characteristics used**	**Potentially available**
Grieco et al. [2024] US autos, 1980–2018	23 (Footprint, Van, Truck, SUV etc)	dozens or hundreds (factory installed features)
Decarolis et al. [2020] Medicare	9 (Deductible, gap coverage, enhanced, number of drugs etc)	hundreds (Formulary details)
Berry and Jia [2010] US airlines	10 (Connections, destinations, No. departures, distance etc)	hundreds (Plane size, seats, hub, transit airports, departure time etc)
Numerator Juice	–	hundreds (brand, flavour, holding company, trading company etc.)

Contributions

1. A feasible Neyman orthogonal score for the BLP model with high-dimensional product characteristics
   - Generic orthogonalization inverts high-dimensional matrices: requires high-level assumptions without economic interpretation
   - Instead: exploit the BLP share inversion at the true nonlinear parameter values to orthogonalize via residualized instruments
   - $\sqrt{T}$-asymptotic normality for a low-dimensional subset of parameters

Contributions

1. A feasible Neyman orthogonal score for the BLP model with high-dimensional product characteristics
   - Generic orthogonalization inverts high-dimensional matrices: requires high-level assumptions without economic interpretation
   - Instead: exploit the BLP share inversion at the true nonlinear parameter values to orthogonalize via residualized instruments
   - $\sqrt{T}$-asymptotic normality for a low-dimensional subset of parameters

2. Estimation of nuisance parameters under approximate sparsity when the nonlinear parameter is unknown
   - Mean utility is recovered from shares only at the true nonlinear parameter, which is itself unknown
   - Preliminary estimator of the nonlinear parameter, with error propagation and rate results

Literature

Nonparametric BLP models

- Nonparametric RC distribution, inverse demand or choice function, building on Berry and Haile [2014]: Dunker et al. [2023], Lu et al. [2023], Wang [2023], Compiani [2022]
- **This paper:** the reverse, with parametric RC and high-dimensional characteristics

High-dimensional BLP models

- BLP model: Berry et al. [1995] and Berry et al. [2004]
- High-dimensional characteristics in BLP: Gillen et al. [2019], Sawada and Kawaguchi [2020]
- **This paper:** fixed dimensional nonlinear parameters + nonlinear nuisance functions $\implies$ valid inference

High-dimensional inference

- Neyman orthogonalisation and regularisation: Chernozhukov et al. [2018], Belloni et al. [2018]
- **This paper:** apply to the BLP setting.

Outline

Standard BLP Model

BLP with High-Dimensional Controls: Neyman Orthogonal Estimation

BLP with High-Dimensional Controls: Approximate Sparsity

Simulation

Empirical Application: Chicago Tuna Market

Standard BLP Model

- T markets, J products per market, outside good 0
- Utility of consumer i for product j in market t:

$$u_{ijt} = \underbrace{x_{jt}'\beta_0 + \alpha_0 p_{jt} + \xi_{jt}}_{\text{mean utility, common to all } i} + b_i p_{jt} + \epsilon_{ijt}, \qquad u_{i0t} = \epsilon_{i0t}$$

  - Observed: x_{jt}, p_{jt}, shares s_{jt}; unobserved: ξ_{jt}
  - $\alpha_0 + b_i$: price sensitivity, $b_i \sim F(\cdot, \sigma_0)$; ϵ_{ijt} i.i.d. type I extreme value
- Price is endogenous, $\mathbb{E}[\xi_{jt} \mid p_{jt}] \neq 0$; instruments: $\mathbb{E}[\xi_{jt} \mid x_{jt}, z_{jt}] = 0$
- Parameters: $\theta_0 = (\sigma_0, \alpha_0, \beta_0')'$

From Market Shares to Mean Utility

- Share function: for any $\delta = (\delta_1, \ldots, \delta_J)' \in \mathbb{R}^J$ and any σ,

$$s_j(\delta; p_t, \sigma) = \int \frac{\exp(\delta_j + p_{jt} b)}{1 + \sum_{k=1}^{J} \exp(\delta_k + p_{kt} b)} \, dF(b, \sigma)$$

- Observed shares: $s_t = s(\delta_t^0; p_t, \sigma_0)$, where $\delta_{jt}^0 = x_{jt}' \beta_0 + \alpha_0 p_{jt} + \xi_{jt}$
- $s(\cdot\,; p_t, \sigma)$ is invertible for every σ [Berry et al., 1995]. Define

$$y_t(\sigma) = \big(y_{1t}(\sigma), \ldots, y_{Jt}(\sigma)\big)' := s^{-1}(s_t; p_t, \sigma)$$

  computable from data (s_t, p_t) for any value σ
- At the true value:

$$\boxed{y_{jt}(\sigma_0) = x_{jt}' \beta_0 + \alpha_0 p_{jt} + \xi_{jt}}$$

Standard BLP Estimation

1. Given σ, invert shares: $y_t(\sigma) = s^{-1}(s_t; p_t, \sigma)$
2. Recover the unobservable:

$$\xi_{jt}(\theta) = y_{jt}(\sigma) - x_{jt}'\beta - \alpha p_{jt}, \qquad \xi_{jt}(\theta_0) = \xi_{jt}$$

3. GMM with $\tilde{z}_{jt} = (z_{jt}', x_{jt}')'$:

$$\hat{\theta} = \arg\min_{\theta} \; \hat{g}(\theta)' W_T \, \hat{g}(\theta), \qquad \hat{g}(\theta) = \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \tilde{z}_{jt} \, \xi_{jt}(\theta)$$

Outline

Standard BLP Model

BLP with High-Dimensional Controls: Neyman Orthogonal Estimation

BLP with High-Dimensional Controls: Approximate Sparsity

Simulation

Empirical Application: Chicago Tuna Market

BLP models with high-dimensional controls

We generalize the BLP model to accommodate potentially high-dimensional characteristics:

- Utility:

$$u_{ijt} = f_{u0}(x_{jt}) + \alpha_0 p_{jt} + \xi_{jt} + b_i p_{jt} + \epsilon_{ijt}$$

  - f_{u0} is unknown and x_{jt} is potentially high-dimensional

- Instrument:

$$\mathbb{E}[\xi_{jt} \mid x_{jt}, z_{jt}] = 0, \qquad z_{jt} = f_{z0}(x_{jt}) + \epsilon^z_{jt}, \quad \mathbb{E}[\epsilon^z_{jt} | x_{jt}] = 0.$$

- Share inversion: $y_{jt}(\sigma_0) = f_{u0}(x_{jt}) + \alpha_0 p_{jt} + \xi_{jt}$
- Parameters of interest: $\theta_{10} = (\sigma_0, \alpha_0)$; nuisance functions: f_{u0}, f_{z0}

Why Standard Approaches Fail

- **Unpenalized GMM** on $\mathbb{E}\left[(z_{jt}', x_{jt}')' \xi_{jt}(\theta)\right] = 0$:
  - **Imprecise:** $d_x + 2$ parameters and $d_x + d_z$ moments; many irrelevant controls inflate variance
  - **Infeasible:** once $d_x + 2 > JT$
- **Regularize, then plug in**
  - First-step error enters the moment at first order $\Rightarrow$ Bias vanishes slower than $T^{-1/2}$ $\Rightarrow$ $\sqrt{T}(\hat{\theta}_1 - \theta_{10})$ not centred at 0[Belloni et al., 2018, Chernozhukov et al., 2018]

$\Rightarrow$ Need a moment that is insensitive to first-step errors

Neyman Orthogonal Moment

- Moment $\mathbb{E}[z_{jt}\xi_{jt}] = 0$
- Residualize *both* terms:

$$\psi(w_t, \theta_1, f_z, f_u) = \tfrac{1}{J}\sum_{j=1}^{J} \underbrace{\big(z_{jt} - f_z(x_{jt})\big)}_{\text{instrument net of controls}} \underbrace{\big(y_{jt}(\sigma) - \alpha p_{jt} - f_u(x_{jt})\big)}_{\xi}$$

  - At the truth: $\mathbb{E}\psi(w_t, \theta_{10}, f_{z0}, f_{u0}) = \mathbb{E}\,\tfrac{1}{J}\sum_{j=1}^{J}[\epsilon^z_{jt}\xi_{jt}] = 0$
  - insensitive to first-step errors
- **Intuition:** IV version of Frisch–Waugh–Lovell
- **Special cases:**
  - with σ_0 known, partially linear IV score [Chernozhukov et al., 2018]; BLP adds σ via $y_{jt}(\sigma)$
  - if f_{u0}, f_{z0} linear, the efficient orthogonal score [Chernozhukov et al., 2018] equals $G\Omega^{-1}\psi$

Details

Why It Works: Errors Enter Only as a Product

- Write $f_z(x_{jt}) = f_{z0} + \Delta_z$, $f_u(x_{jt}) = f_{u0} + \Delta_u$
  - Dependence on x_{jt} suppressed for clarity
- Consider $\mathbb{E}\psi(w_t, \theta_{10}, f_z, f_u) - \mathbb{E}\psi(w_t, \theta_{10}, f_{z0}, f_{u0})$. Each product term of ψ:

$$\begin{aligned}\big(z_{jt} - f_z\big)\big(y_{jt}(\sigma_0) - \alpha_0 p_{jt} - f_u\big) &= \big(\epsilon^z_{jt} - \Delta_z\big)\big(\xi_{jt} - \Delta_u\big) \\ &= \epsilon^z_{jt}\xi_{jt} - \Delta_z\,\xi_{jt} - \epsilon^z_{jt}\,\Delta_u + \underbrace{\Delta_z\,\Delta_u}_{\text{error}\times\text{error}} \end{aligned} \tag{1}$$

Why It Works: Errors Enter Only as a Product

- Write $f_z(x_{jt}) = f_{z0} + \Delta_z$, $f_u(x_{jt}) = f_{u0} + \Delta_u$
  - Dependence on x_{jt} suppressed for clarity
- Consider $\mathbb{E}\psi(w_t, \theta_{10}, f_z, f_u) - \mathbb{E}\psi(w_t, \theta_{10}, f_{z0}, f_{u0})$. Each product term of ψ:

$$\begin{aligned}\big(z_{jt} - f_z\big)\big(y_{jt}(\sigma_0) - \alpha_0 p_{jt} - f_u\big) &= \big(\epsilon^z_{jt} - \Delta_z\big)\big(\xi_{jt} - \Delta_u\big) \\ &= \epsilon^z_{jt}\xi_{jt} - \Delta_z\,\xi_{jt} - \epsilon^z_{jt}\,\Delta_u + \underbrace{\Delta_z\,\Delta_u}_{\text{error}\times\text{error}} \end{aligned} \tag{1}$$

- **Middle terms vanish:** if $\Delta_z(x_{jt})$, $\Delta_u(x_{jt})$ deterministic functions of x_{jt}, $\mathbb{E}[\xi_{jt} \mid x_{jt}, z_{jt}] = 0$, $\mathbb{E}[\epsilon^z_{jt} \mid x_{jt}] = 0 \Rightarrow$ mean zero

Why It Works: Errors Enter Only as a Product

- Write $f_z(x_{jt}) = f_{z0} + \Delta_z$, $f_u(x_{jt}) = f_{u0} + \Delta_u$
  - Dependence on x_{jt} suppressed for clarity
- Consider $\mathbb{E}\psi(w_t, \theta_{10}, f_z, f_u) - \mathbb{E}\psi(w_t, \theta_{10}, f_{z0}, f_{u0})$. Each product term of ψ:

$$\begin{aligned}\big(z_{jt} - f_z\big)\big(y_{jt}(\sigma_0) - \alpha_0 p_{jt} - f_u\big) &= \big(\epsilon^z_{jt} - \Delta_z\big)\big(\xi_{jt} - \Delta_u\big) \\ &= \epsilon^z_{jt}\xi_{jt} - \Delta_z\,\xi_{jt} - \epsilon^z_{jt}\,\Delta_u + \underbrace{\Delta_z\,\Delta_u}_{\text{error}\times\text{error}} \end{aligned} \tag{1}$$

- **Middle terms vanish:** if $\Delta_z(x_{jt})$, $\Delta_u(x_{jt})$ deterministic functions of x_{jt}, $\mathbb{E}[\xi_{jt} \mid x_{jt}, z_{jt}] = 0$, $\mathbb{E}[\epsilon^z_{jt} \mid x_{jt}] = 0 \Rightarrow$ mean zero
- **Remaining term is second order:** $\|\mathbb{E}[\Delta_z \Delta_u]\| \leq \|\Delta_z\|_{L^2}\|\Delta_u\|_{L^2} = o(T^{-\frac{1}{4}}) \cdot o(T^{-\frac{1}{4}}) = o(T^{-\frac{1}{2}}) \Rightarrow$ negligible

Why It Works: Errors Enter Only as a Product

- Write $f_z(x_{jt}) = f_{z0} + \Delta_z$, $f_u(x_{jt}) = f_{u0} + \Delta_u$
  - Dependence on x_{jt} suppressed for clarity
- Consider $\mathbb{E}\psi(w_t, \theta_{10}, f_z, f_u) - \mathbb{E}\psi(w_t, \theta_{10}, f_{z0}, f_{u0})$. Each product term of ψ:

$$\begin{aligned}\big(z_{jt} - f_z\big)\big(y_{jt}(\sigma_0) - \alpha_0 p_{jt} - f_u\big) &= \big(\epsilon^z_{jt} - \Delta_z\big)\big(\xi_{jt} - \Delta_u\big) \\ &= \epsilon^z_{jt}\xi_{jt} - \Delta_z\,\xi_{jt} - \epsilon^z_{jt}\,\Delta_u + \underbrace{\Delta_z\,\Delta_u}_{\text{error}\times\text{error}} \end{aligned} \tag{1}$$

- **Middle terms vanish:** if $\Delta_z(x_{jt})$, $\Delta_u(x_{jt})$ deterministic functions of x_{jt}, $\mathbb{E}[\xi_{jt} \mid x_{jt}, z_{jt}] = 0$, $\mathbb{E}[\epsilon^z_{jt} \mid x_{jt}] = 0 \Rightarrow$ mean zero
- **Remaining term is second order:** $\|\mathbb{E}[\Delta_z \Delta_u]\| \leq \|\Delta_z\|_{L^2}\|\Delta_u\|_{L^2} = o(T^{-\frac{1}{4}}) \cdot o(T^{-\frac{1}{4}}) = o(T^{-\frac{1}{2}}) \Rightarrow$ negligible

Why It Works: Errors Enter Only as a Product

- Write $f_z(x_{jt}) = f_{z0} + \Delta_z$, $f_u(x_{jt}) = f_{u0} + \Delta_u$
  - Dependence on x_{jt} suppressed for clarity
- Consider $\mathbb{E}\psi(w_t, \theta_{10}, f_z, f_u) - \mathbb{E}\psi(w_t, \theta_{10}, f_{z0}, f_{u0})$. Each product term of ψ:

$$\begin{aligned}(z_{jt} - f_z)\big(y_{jt}(\sigma_0) - \alpha_0 p_{jt} - f_u\big) &= \big(\epsilon^z_{jt} - \Delta_z\big)\big(\xi_{jt} - \Delta_u\big) \\ &= \epsilon^z_{jt}\xi_{jt} - \Delta_z\,\xi_{jt} - \epsilon^z_{jt}\,\Delta_u + \underbrace{\Delta_z\,\Delta_u}_{\text{error}\times\text{error}} \qquad (1)\end{aligned}$$

- **Middle terms vanish:** if $\Delta_z(x_{jt})$, $\Delta_u(x_{jt})$ deterministic functions of x_{jt}, $\mathbb{E}[\xi_{jt} \mid x_{jt}, z_{jt}] = 0$, $\mathbb{E}[\epsilon^z_{jt} \mid x_{jt}] = 0 \Rightarrow$ mean zero
- **Remaining term is second order:** $\|\mathbb{E}[\Delta_z \Delta_u]\| \leq \|\Delta_z\|_{L^2}\|\Delta_u\|_{L^2} = o(T^{-\frac{1}{4}}) \cdot o(T^{-\frac{1}{4}}) = o(T^{-\frac{1}{2}}) \Rightarrow$ negligible
- Sample average of (1) over markets, with f_u and f_z estimated from the data $\Rightarrow$ Δ_u and Δ_z not deterministic $\Rightarrow$ sample splitting

Why Sample Splitting?

- **Problem:** middle terms may not vanish if Δ_z, Δ_u are *estimated* functions
  - e.g. $\hat{f}_u$ fitted on the full sample $\Rightarrow \Delta_u(x_{jt}) = \hat{f}_u(x_{jt}) - f_{u0}(x_{jt})$ correlated with ϵ^z_{jt} in sample $\Rightarrow \frac{1}{T}\sum_t \epsilon^z_{jt}\Delta_u(x_{jt})$ no longer averages to zero

Why Sample Splitting?

- **Problem:** middle terms may not vanish if Δ_z, Δ_u are *estimated* functions
  - e.g. $\hat{f}_u$ fitted on the full sample $\Rightarrow \Delta_u(x_{jt}) = \hat{f}_u(x_{jt}) - f_{u0}(x_{jt})$ correlated with ϵ^z_{jt} in sample $\Rightarrow \frac{1}{T}\sum_t \epsilon^z_{jt}\Delta_u(x_{jt})$ no longer averages to zero
- **Fix:** estimate $\hat{f}_u$ on *one* subsample, evaluate ψ on *another*
  - given the subsample used to estimate $\hat{f}_u$, Δ_u is a deterministic function $\Rightarrow \epsilon^z_{jt}\Delta_u(x_{jt})$ is an average of independent mean-zero terms: $O_p\big(\|\Delta_u\|_{L^2}/\sqrt{T}\big) = o_p(T^{-\frac{1}{2}})$

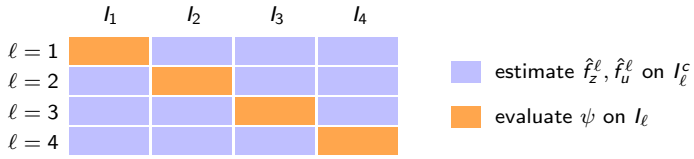
Why Sample Splitting?

- **Problem:** middle terms may not vanish if Δ_z, Δ_u are *estimated* functions
  - e.g. $\hat{f}_u$ fitted on the full sample $\Rightarrow \Delta_u(x_{jt}) = \hat{f}_u(x_{jt}) - f_{u0}(x_{jt})$ correlated with ϵ^z_{jt} in sample $\Rightarrow \frac{1}{T}\sum_t \epsilon^z_{jt}\Delta_u(x_{jt})$ no longer averages to zero
- **Fix:** estimate $\hat{f}_u$ on *one* subsample, evaluate ψ on *another*
  - given the subsample used to estimate $\hat{f}_u$, Δ_u is a deterministic function $\Rightarrow \epsilon^z_{jt}\Delta_u(x_{jt})$ is an average of independent mean-zero terms: $O_p\left(\|\Delta_u\|_{L^2}/\sqrt{T}\right) = o_p(T^{-\frac{1}{2}})$
- **Intuition:** as in machine learning, never evaluate a model on the data used to fit it

Why Sample Splitting?

- **Problem:** middle terms may not vanish if Δ_z, Δ_u are *estimated* functions
  - e.g. $\hat{f}_u$ fitted on the full sample $\Rightarrow \Delta_u(x_{jt}) = \hat{f}_u(x_{jt}) - f_{u0}(x_{jt})$ correlated with ϵ^z_{jt} in sample $\Rightarrow \frac{1}{T}\sum_t \epsilon^z_{jt}\Delta_u(x_{jt})$ no longer averages to zero
- **Fix:** estimate $\hat{f}_u$ on *one* subsample, evaluate ψ on *another*
  - given the subsample used to estimate $\hat{f}_u$, Δ_u is a deterministic function $\Rightarrow \epsilon^z_{jt}\Delta_u(x_{jt})$ is an average of independent mean-zero terms: $O_p\left(\|\Delta_u\|_{L^2}/\sqrt{T}\right) = o_p(T^{-\frac{1}{2}})$
- **Intuition:** as in machine learning, never evaluate a model on the data used to fit it
- **Alternative solution:** impose stronger complexity conditions on the set of feasible $\hat{f}$, e.g. stronger sparsity condition for LASSO [Chen and Chiang, 2026]

Sample Splitting Method: Cross-Fitting



Construct the cross-fitted moment

1. For each ℓ, estimate $\hat{f}_z^\ell, \hat{f}_u^\ell$ (blue)
2. Evaluate $\psi(w_t, \theta_1, \hat{f}_z^\ell, \hat{f}_u^\ell)$ for $t \in I_\ell$ (orange)
3. Rotate over $\ell = 1, \ldots, L$ and average:

$$\bar{\psi}(\theta_1) = \frac{1}{T} \sum_{\ell=1}^{L} \sum_{t \in I_\ell} \psi(w_t, \theta_1, \hat{f}_z^\ell, \hat{f}_u^\ell)$$

Estimate $\check{\theta}_1$ by GMM

$$\check{\theta}_1 = \underset{\theta_1 \in \Theta_\sigma \times \Theta_\alpha}{\arg\min}\ \bar{\psi}(\theta_1)' \, \hat{W} \, \bar{\psi}(\theta_1)$$

Assumption: Rate Condition

Condition 1

For each $\ell = 1, \ldots, L$*:*

$$\|\hat{f}_z^\ell - f_{z0}\|_{L^2} = o_p(T^{-\frac{1}{4}}), \qquad \|\hat{f}_u^\ell - f_{u0}\|_{L^2} = o_p(T^{-\frac{1}{4}})$$

- **Why** $T^{-1/4}$**:** only error $\times$ error survives $\Rightarrow \|\Delta_z\|_{L^2}\|\Delta_u\|_{L^2} = o_p(T^{-\frac{1}{2}}) \Rightarrow$ negligible $\Rightarrow$ room for regularization / machine learning
- **Why the** L^2 **norm:** $\|\hat{f}^\ell - f_0\|_{L^2}^2 = \mathbb{E}_x\|\hat{f}^\ell(x) - f_0(x)\|_2^2$ over a new draw x, with $\hat{f}^\ell$ held fixed $\Rightarrow$ well defined because $\hat{f}^\ell$ is estimated on I_ℓ^c (cross-fitting)
- **Any learner:** with cross-fitting, only the rate matters, e.g. LASSO under approximate sparsity

Assumption: Identification and Rank Conditions

- **Identification** (global) $\Rightarrow$ consistency — Condition
  - $\mathbb{E}\psi(w_t, \theta_1, f_{z0}, f_{u0}) = 0$ only at θ_{10}
  - Linear IV analogue: $\mathbb{E}[z(y - \alpha p)] = 0$ only at α_0
- **Rank** (local) $\Rightarrow$ asymptotic normality — Condition
  - $G = \mathbb{E}_J \mathbb{E}\left[(\partial_\sigma y_{jt}(\sigma_0)', -p_{jt})' \epsilon_{jt}^{z\prime}\right]$ has full rank
  - Residualized instrument must move price (α) and $\partial_\sigma y_{jt}$ (σ) $\Rightarrow$ $d_z \geq 2$
  - Linear IV analogue: $\mathrm{Cov}(\epsilon_{jt}^z, p_{jt}) \neq 0$, the first stage after partialling out x_{jt}
- **Both uniform in** T**:** (triangular array) f_{u0}, f_{z0} and the data distribution vary with T

Theory: Main Results

Theorem 1 (Consistency)

Under regularity Cond., *identification* Cond. *and Condition 1:* $\check{\theta}_1 \xrightarrow{p} \theta_{10}$

Theorem 2 (Asymptotic Normality)

If, in addition, the rank condition holds Cond.*:*

$$P^{-1/2}\sqrt{T}(\check{\theta}_1 - \theta_{10}) \xrightarrow{d} N(0, I_{d_{\theta_1}}), \qquad P = (GWG')^{-1}GW\Omega WG'(GWG')^{-1}$$

where $\Omega = \text{Var}\left(\frac{1}{J}\sum_{j=1}^{J}\epsilon_{jt}^{z}\xi_{jt}\right)$.

- $W = \Omega^{-1} \Rightarrow P = (G\Omega^{-1}G')^{-1}$, the efficient GMM variance
- Extends to ratios such as $\check{\alpha}/\check{\beta}_k$; elasticities consistently estimated

Outline

Standard BLP Model

BLP with High-Dimensional Controls: Neyman Orthogonal Estimation

BLP with High-Dimensional Controls: Approximate Sparsity

Simulation

Empirical Application: Chicago Tuna Market

Approximate Sparsity

Why: f lies in an inf.-dim. function space $\Rightarrow$ approximated by a fin.-dim. function space.
Idea: many candidate controls, few relevant, unknown which $\Rightarrow$ let LASSO find them

Nuisance	**Sparse approximation** (at most d_0 nonzeros)
$f_{u0}(x_{jt})$	$\beta_0' \tilde{x}_{jt}$
$f_{z0}(x_{jt})$	$\Pi_0 \tilde{x}_{jt}$
$f_{p0}(x_{jt}, z_{jt})$	$\beta_{pz0}' z_{jt} + \beta_{px0}' \tilde{x}_{jt}$

- **Dictionary:** $\tilde{x} = x$ or basis functions
- **Approximate, not exact:** approximation error $\lesssim d_0/T$
- **Selection:** ad hoc selection: a few by judgment; PCA: a few components; ours: by data
- **Auxiliary f_{p0}:** $p_{jt} = f_{p0}(x_{jt}, z_{jt}) + \epsilon_{jt}^p$ e.g. $\mathbb{E}[p_{jt} \mid x_{jt}, z_{jt}]$; does not enter NO moment

Estimating the Nuisance Functions: Roadmap

Need: $\hat{f}_z$ and $\hat{f}_u$ at rate $o_p(T^{-1/4})$ For each fold ℓ, using only markets in I_ℓ^c:

1. **Step 1** (LASSO):
   - z_{jt} on $\tilde{x}_{jt} \Rightarrow \hat{f}_z^\ell$
   - p_{jt} on $(z_{jt}, \tilde{x}_{jt}) \Rightarrow$ predicted price $\hat{f}_p^\ell$
2. **Step 2A** (preliminary nonlinear parameter σ):
   - penalized minimum distance on the moments $\Rightarrow \tilde{\sigma}^\ell$
3. **Step 2B** (LASSO):
   - $y_{jt}(\tilde{\sigma}^\ell)$ on $\hat{f}_p^\ell$ and $\tilde{x}_{jt} \Rightarrow \hat{f}_u^\ell$

Then: $(\hat{f}_z^\ell, \hat{f}_u^\ell)$ enter the cross-fitted orthogonal moment $\Rightarrow \breve{\theta}_1$

Step 1: $\hat{f}_z$ and $\hat{f}_p$

Regression of instruments on controls, for each $i = 1, \ldots, d_z$:

$$\hat{\Pi}_i^\ell = \arg\min_{\Pi_i} \frac{1}{J|I_\ell^c|} \sum_{j, I_\ell^c} (z_{jt,i} - \tilde{x}_{jt}' \Pi_i)^2 + \lambda_\Pi \|\Pi_i\|_1$$

Regression of price on instruments and controls:

$$(\hat{\beta}_{pz}^\ell, \hat{\beta}_{px}^\ell) = \arg\min_{\beta_{pz}, \beta_{px}} \frac{1}{J|I_\ell^c|} \sum_{j, I_\ell^c} (p_{jt} - z_{jt}' \beta_{pz} - \tilde{x}_{jt}' \beta_{px})^2 + \lambda_p \|(\beta_{pz}', \beta_{px}')\|_1$$

Fitted Nuisance Functions: $\hat{f}_z^\ell(x) = \hat{\Pi}^\ell \tilde{x}; \quad \hat{f}_p^\ell(x, z) = z' \hat{\beta}_{pz}^\ell + \tilde{x}' \hat{\beta}_{px}^\ell$

Step 2A: Preliminary Estimator of nonlinear parameter σ

Full parameter $\theta = (\sigma', \alpha, \beta')'$. Moments on I_ℓ^c, with $\tilde{z}_{jt} = (z'_{jt}, \tilde{x}'_{jt})'$:

$$\hat{g}^\ell(\theta) = \frac{1}{J|I_\ell^c|} \sum_{j, I_\ell^c} \tilde{z}_{jt}\, \xi_{jt}(\theta), \qquad \xi_{jt}(\theta) = y_{jt}(\sigma) - \alpha p_{jt} - \tilde{x}'_{jt}\beta$$

$$(\tilde{\sigma}^\ell, \tilde{\alpha}^\ell, \tilde{\beta}^\ell) = \arg \min_{\sigma \in \Theta_\sigma,\, \alpha \in \Theta_\alpha,\, \beta} \left\{ \|\hat{g}^\ell(\theta)\|_\infty + \lambda_{\tilde{\theta}} \|\beta\|_1 \right\}$$

- ℓ_∞**, not** ℓ_2**:** sampling error is $\sqrt{d_{\tilde{z}}/T}$ in ℓ_2 but only $\sqrt{\log d_{\tilde{z}}/T}$ in $\ell_\infty \Rightarrow$ $d_{\tilde{z}}$ may exceed T means $\sqrt{d_{\tilde{z}}/T} \nrightarrow 0$
- **Only** β **penalized:** penalizing σ or α would shrink the parameters this step must estimate accurately
- **Computation:** for fixed σ, a linear program in $(\alpha, \beta) \Rightarrow$ grid search over low-dimensional σ

Step 2B: LASSO for $\hat{f}_u$

$$(\hat{\alpha}^\ell, \hat{\beta}^\ell) = \arg\min_{\alpha,\beta} \frac{1}{J|I_\ell^c|} \sum_{j,I_\ell^c} \left(y_{jt}(\tilde{\sigma}^\ell) - \alpha \hat{f}_p^\ell(x_{jt}, z_{jt}) - \tilde{x}_{jt}'\beta\right)^2 + \lambda_\beta \|\beta\|_1$$

- **Given approximate sparsity and $\tilde{\sigma}^\ell$:** $y_{jt}(\tilde{\sigma}^\ell)$ is approximately linear in $(p_{jt}, \tilde{x}_{jt}) \Rightarrow$ LASSO
- **Predicted price:** p_{jt} endogenous $\Rightarrow$ replaced by $\hat{f}_p^\ell$, as in 2SLS
- **Error propagation:** error in $\tilde{\sigma}^\ell$ passes through the share inversion into $y_{jt}(\tilde{\sigma}^\ell)$; $\tilde{\sigma}^\ell$ fast enough $\Rightarrow$ $\hat{f}_u^\ell$ attains $o_p(T^{-1/4})$

Fitted Nuisance Functions: $\hat{f}_u^\ell(x) = \tilde{x}'\hat{\beta}^\ell$

Theory: Key Conditions for the Nuisance Estimators

- **Approximate sparsity** Details
  - Each nuisance can be approximated by at most d_0 terms; approximation error $\lesssim d_0/T$
  - Sparsity index small: $\sqrt{d_0 \log(d_x \vee T)/T} = o(T^{-1/4})$
- **Sparse eigenvalues** Details
  - Gram matrices of $(f_{p0}(x_{jt}, z_{jt}), x'_{jt})$ and (z'_{jt}, x'_{jt}) well conditioned on every small subset of regressors
  - can hold with $d_x > T$
  - Fails if
    - there are exact duplicates (all product dummies plus intercept),
    - instruments carry no information beyond a small set of x_{jt}

Theory: Nuisance Convergence Rates

Theorem 3 (Nuisance Convergence Rates)

Under the conditions above, regularity conditions Appendix *and suitably chosen tuning parameters* λ's*, for each* $l = 1, \ldots, L$*, w.p.a.1,*

$$\|\hat{f}_z^\ell - f_{z0}\|_{L^2} + \|\hat{f}_u^\ell - f_{u0}\|_{L^2} \lesssim \sqrt{\frac{d_0 \log(d_x \vee T)}{T}} = o(T^{-1/4})$$

Theory: Nuisance Convergence Rates

Theorem 3 (Nuisance Convergence Rates)

Under the conditions above, regularity conditions Appendix *and suitably chosen tuning parameters* λ's*, for each* $l = 1, \ldots, L$*, w.p.a.1,*

$$\|\hat{f}_z^\ell - f_{z0}\|_{L^2} + \|\hat{f}_u^\ell - f_{u0}\|_{L^2} \lesssim \sqrt{\frac{d_0 \log(d_x \vee T)}{T}} = o(T^{-1/4})$$

- Rate
  - Oracle $\sqrt{d_0/T}$
  - Price of not knowing which d_0 terms matter: $\sqrt{\log(d_x \vee T)}$
- $\Rightarrow$ Condition 1 holds $\Rightarrow$ with Theorem 2, $\check{\theta}_1$ is asymptotically normal with LASSO nuisance estimates

Theory: Nuisance Convergence Rates

Theorem 3 (Nuisance Convergence Rates)

Under the conditions above, regularity conditions Appendix *and suitably chosen tuning parameters* λ's*, for each* $l = 1, \ldots, L$*, w.p.a.1,*

$$\|\hat{f}_z^\ell - f_{z0}\|_{L^2} + \|\hat{f}_u^\ell - f_{u0}\|_{L^2} \lesssim \sqrt{\frac{d_0 \log(d_x \vee T)}{T}} = o(T^{-1/4})$$

- Rate
  - Oracle $\sqrt{d_0/T}$
  - Price of not knowing which d_0 terms matter: $\sqrt{\log(d_x \vee T)}$
- $\Rightarrow$ Condition 1 holds $\Rightarrow$ with Theorem 2, $\check{\theta}_1$ is asymptotically normal with LASSO nuisance estimates

Takeaway: both preliminary estimators converge fast enough ($o(T^{-1/4})$).

Outline

Standard BLP Model

BLP with High-Dimensional Controls: Neyman Orthogonal Estimation

BLP with High-Dimensional Controls: Approximate Sparsity

Simulation

Empirical Application: Chicago Tuna Market

Simulation Design I

- **Data generating process:**

$$y_{jt} = \alpha_0 p_{jt} + x'_{jt}\beta_0 + \xi_{jt}, \quad z_{jt} = \Pi_0 x_{jt} + \epsilon^z_{jt}, \quad p_{jt} = x'_{jt}\beta_{px0} + \epsilon^p_{jt}$$

- **Characteristics:** $\dim(x_{jt}) = 300$
- **Markets and goods:** $T = 50$; $J = 4$
- **Instruments:** $\dim(z_{jt}) = 12$
- **Endogeneity:** ϵ^p_{jt} correlated with ξ_{jt} and ϵ^z_{jt}

Simulation Design II

Implementation:

- Sample splitting: $L = 10$ folds of market indices
- Tuning parameters selected by cross-validation
- Inference via t-tests

Simulation Results for σ

Table: Simulation Results for σ

	Bias	SD	RMSE	Rej. rate
GMM optimal IV	0.017	0.186	0.186	0.010
Preliminary	−0.113	0.110	0.157	0.520
No Neyman orthogonal	−0.102	0.089	0.135	0.090
Neyman orthogonal	−0.021	0.120	0.121	0.040
Oracle 1	0.003	0.077	0.077	0.030
Oracle 2	−0.004	0.093	0.093	0.020

100 Monte Carlo replications. Rej. rate at 5% level.

Simulation Results for α

Table: Simulation Results for α

	Bias	SD	RMSE	Rej. rate
GMM optimal IV	-0.016	0.140	0.141	0.040
Preliminary	0.093	0.081	0.123	0.790
No Neyman orthogonal	0.116	0.068	0.135	0.670
Neyman orthogonal	0.030	0.069	0.075	0.070
Oracle 1	0.001	0.032	0.032	0.040
Oracle 2	0.005	0.051	0.051	0.040

100 Monte Carlo replications. Rej. rate at 5% level.

Outline

Standard BLP Model

BLP with High-Dimensional Controls: Neyman Orthogonal Estimation

BLP with High-Dimensional Controls: Approximate Sparsity

Simulation

Empirical Application: Chicago Tuna Market

Data: Dominick's Finer Foods

- **Sample:** Weekly store-level tuna scanner data, 1989–1995
  - 86 Chicago-area stores
- **Markets:** Store $\times$ four-week periods
  - 6,138 markets after excluding 0 share markets; shares measured as units per purchasing visit
- **Choice set:** Eight leading product groups plus *Other tuna*
  - 9 alternatives, fixed across markets
- **Controls:** Product-specific neighbourhood demographics and product, year, month fixed effects
  - 456 candidate regressors; 412 retained

Estimation Results: Demand Parameters

$$\alpha_i = -\exp(\mu + \sigma\nu_i), \qquad \nu_i \sim N(0,1).$$

	Random-coefficient logit			Logit
	NO	GMM (Unpenalized)	GMM (optimal IV)	2SLS
$\hat{\mu}$	2.034	3.368	3.552	–
Std. error	(0.967)	–	–	–
$\hat{\sigma}$	0.700	1.011	1.923	–
Std. error	(0.435)	–	–	–
$\hat{\alpha}$	–	–	–	-0.063
Std. error	–	–	–	(0.141)
Step A penalty	0.06	–	–	–
Alternatives		9		
Observations		55,242		
Markets		6,138		

Notes: Reported standard errors are clustered by market. GMM standard errors are omitted due to numerically rank-deficient Jacobians.

Own-Price Elasticities

- NO elasticities are tightly concentrated in economically plausible magnitudes
- Two GMM estimators produce a substantially heavier left tail

	Random-coefficient logit			Logit
Statistic	NO	GMM (Unpenalized)	GMM (optimal IV)	2SLS
Mean	−2.629	−4.005	−1.710	−0.057
Std. deviation	0.375	1.186	0.592	0.020
Minimum	−4.461	−13.549	−16.952	−0.120
1st quartile	−2.819	−4.663	−2.052	−0.060
Median	−2.556	−3.684	−1.675	−0.049
3rd quartile	−2.374	−3.169	−1.322	−0.044
Maximum	−1.213	−0.820	−0.436	−0.016
Alternatives		9		
Observations		55,242		

Held-Out Market Prediction

Estimator	MSE	Relative MSE	RMSE	MAE	Mean s	Mean $\widehat{s}$
NO	0.000452	0.126	0.02125	0.00827	0.00723	0.01015
GMM (unpenalized)	0.003576	1.000	0.05980	0.01718	0.00723	0.01767
GMM (optimal IV)	0.003740	1.046	0.06116	0.01726	0.00723	0.01758

Backup: Connection to the Efficient Orthogonal Score

Suppose $f_{u0}(x) = x'\beta_0$, $f_{z0}(x) = \Pi_0 x$. Full GMM moment with $\tilde{z}_{jt} = (z'_{jt}, x'_{jt})'$:

$$m(w_t, \theta_1, \beta) = \mathbb{E}_J \tilde{z}_{jt}\big(y_{jt}(\sigma) - \alpha p_{jt} - x'_{jt}\beta\big)$$

- Generic efficient orthogonal score [Chernozhukov et al., 2018, eq. 2.18]: $\mu_0 m$,

$$\mu_0 = A'V^{-1} - A'V^{-1}G_\beta(G'_\beta V^{-1} G_\beta)^{-1} G'_\beta V^{-1}$$

  A, G_β: Jacobians of $\mathbb{E}m$ w.r.t. θ_1 and β; $V = \mathrm{Var}(m)$: high-dimensional objects
- **Proposition:** for all θ_1, β, $\mu_0\, m(w_t, \theta_1, \beta) = G\Omega^{-1}\psi(w_t, \theta_1, f_{z0}, x'\beta)$
- Same score up to a fixed $d_{\theta_1} \times d_z$ matrix: with $W = \Omega^{-1}$, attains $(G\Omega^{-1}G')^{-1}$
- ψ needs only two regressions, and stays orthogonal when f_{u0}, f_{z0} are nonlinear

Appendix: Regularity Conditions

1. $(w_t)_{t=1}^T$ i.i.d. and satisfy the utility and instrument equations
2. Density $f(b, \sigma)$ of b_i twice continuously differentiable in σ, with $\int \sup_{\Theta_\sigma} \|\partial_\sigma f\| \, db \leq M$ and $\int \sup_{\Theta_\sigma} \|\partial_\sigma^2 f\| \, db \leq M$
3. $|p_{jt}|$, $\|\epsilon_{jt}^z\|_\infty$, $|\xi_{jt}|$, $|f_{u0}(x_{jt})|$, $\|f_{z0}(x_{jt})\|_\infty \leq M$ a.s.; $\|\theta_{10}\|_1 \leq M$; d_z fixed
4. Θ_σ, Θ_α fixed compact sets, σ_0, α_0 interior
5. $\hat{W}$, W positive definite, $\|\hat{W} - W\|_2 = o_p(1)$, $c \leq \lambda_{\min}(W) \leq \lambda_{\max}(W) \leq C$

- Item 2 makes $y_{jt}(\sigma)$ twice continuously differentiable; holds for normal and lognormal
- Boundedness for simplicity; replaceable by moment conditions

Back

Appendix: Identification Condition

For every $\epsilon > 0$ there is $\eta > 0$, independent of T, such that

$$\inf \left\{ \left\| \mathbb{E}\psi(w_t, \theta_1, f_{z0}, f_{u0}) \right\|_2 : \ \theta_1 \in \Theta_\sigma \times \Theta_\alpha, \ \|\theta_1 - \theta_{10}\|_2 > \epsilon \right\} \geq \eta$$

- Global: θ_{10} is the unique, well-separated zero of the population moment
- Uniform in T: otherwise some $\theta_1 \neq \theta_{10}$ could solve the moment in the limit
- Fixed-dimensional counterparts: Berry et al. [2004], Freyberger [2015]

Back

Appendix: Rank Condition

With $G = \mathbb{E}_J\mathbb{E}\left[(\partial_\sigma y_{jt}(\sigma_0)',\ -p_{jt})'\,\epsilon_{jt}^{z\prime}\right]$ and $\Omega = \text{Var}\left(\frac{1}{J}\sum_j \epsilon_{jt}^z \xi_{jt}\right)$:

1. $\lambda_{\min}(GG') > c > 0$
2. $\lambda_{\min}(\Omega) > c > 0$

- Item 1: local identification; residualized instruments relevant for both σ and α
- Item 2: the orthogonal moment has non-degenerate variance
- Both uniform in T, so G and Ω stay nonsingular in the limit

Back

Appendix: Regularity Conditions for the Nuisance Estimators

In addition to the regularity, identification and rank conditions of Section 2:

1. Price equation: $p_{jt} = f_{p0}(x_{jt}, z_{jt}) + \epsilon^p_{jt}$, $\mathbb{E}[\epsilon^p_{jt}(z'_{jt}, x'_{jt})] = 0$, $\mathbb{E}[\epsilon^p_{jt} f_{p0}(x_{jt}, z_{jt})] = 0$
2. $\lambda_{\max}(\mathbb{E} x_{jt} x'_{jt}) \leq M$
3. $|p_{jt}|$, $\|x_{jt}\|_\infty$, $\|\epsilon^z_{jt}\|_\infty$, $|\xi_{jt}|$, $|f_{u0}(x_{jt})|$, $\|f_{z0}(x_{jt})\|_\infty$, $|f_{p0}(x_{jt}, z_{jt})| \leq M$ a.s.; $\|\theta_{10}\|_1$, $\|\beta_0\|_1$, $\|\Pi_0\|_1 \leq M$; d_z fixed, $d_{\tilde{z}} = O(d_x)$
4. Θ_σ, Θ_α fixed compact sets, σ_0, α_0 interior

- Item 3 strengthens the boundedness condition of Section 2 and replaces it throughout
- For the LASSO steps, replaceable by moment conditions [Belloni et al., 2014]; Step 2A needs sub-Gaussian tails

Back

Appendix: Approximate Sparsity

1. There exist β_0, Π_0, β_{pz0}, β_{px0}, possibly depending on T, with

$$
\begin{aligned}
f_{u0}(x_{jt}) &= x_{jt}'\beta_0 + r_{jt}^u, && \|\beta_0\|_0 \le d_0, && \mathbb{E}(r_{jt}^u)^2 \lesssim d_0/T \\
f_{z0}(x_{jt}) &= \Pi_0 x_{jt} + r_{jt}^z, && \|\Pi_0\|_0 \le d_0, && \mathbb{E}\|r_{jt}^z\|_2^2 \lesssim d_0/T \\
f_{p0}(x_{jt}, z_{jt}) &= z_{jt}'\beta_{pz0} + x_{jt}'\beta_{px0} + r_{jt}^p, && \|(\beta_{pz0}', \beta_{px0}')\|_0 \le d_0, && \mathbb{E}(r_{jt}^p)^2 \lesssim d_0/T
\end{aligned}
$$

2. $\sqrt{d_0 \log(d_x \vee T)/T} = o(T^{-1/4})$

- Approximation errors vanish at the oracle rate $\sqrt{d_0/T}$ $\Rightarrow$ do not dominate the LASSO estimation error

Back

Appendix: Sparse Eigenvalue Condition

For a positive semi-definite M and integer $m \geq 1$:

$$\phi_{\min}(m)[M] = \min_{\delta \neq 0,\ \|\delta\|_0 \leq m} \frac{\delta' M \delta}{\|\delta\|_2^2}, \qquad \phi_{\max}(m)[M] = \max_{\delta \neq 0,\ \|\delta\|_0 \leq m} \frac{\delta' M \delta}{\|\delta\|_2^2}$$

Let $\hat{\Sigma}_\ell^{(1)} = \frac{1}{J|I_\ell^c|} \sum_{j, I_\ell^c} (f_{p0}, x_{jt}')'(f_{p0}, x_{jt}')$ and $\hat{\Sigma}_\ell^{(2)} = \frac{1}{J|I_\ell^c|} \sum_{j, I_\ell^c} \tilde{z}_{jt} \tilde{z}_{jt}'$.

Condition: there exist $\ell_T \to \infty$ and $0 < \underline{\phi} \leq \bar{\phi} < \infty$ such that, w.p.a.1, for all ℓ and $k = 1, 2$,

$$\underline{\phi} \leq \phi_{\min}(\ell_T d_0)\left[\hat{\Sigma}_\ell^{(k)}\right] \leq \phi_{\max}(\ell_T d_0)\left[\hat{\Sigma}_\ell^{(k)}\right] \leq \bar{\phi}$$

- Condition SE of Belloni et al. [2014]; a primitive sufficient condition on population matrices is in the Online Appendix

Back

Appendix: Tuning Parameters

$$\lambda_\Pi = C_\Pi \sqrt{\frac{\log(d_x \vee T)}{T}}, \qquad \lambda_p = C_p \sqrt{\frac{\log(d_x \vee T)}{T}}, \qquad \lambda_\beta = C_\beta \sqrt{\frac{\log(d_x \vee T)}{T}},$$

$$\lambda_{\tilde{\theta}} = C_{\tilde{\theta}} \sqrt{\frac{d_0 \log(d_x \vee T)}{T}}$$

with sufficiently large constants.

- $\lambda_{\tilde{\theta}}$ carries d_0: Step 2A must dominate the approximation error r_{jt}^u in every moment
- In practice: cross-validation (simulation and application)

Back

References I

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2): 608–650, 2014.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, Christian Hansen, and Kengo Kato. High-dimensional econometrics and regularized GMM. *arXiv preprint arXiv:1806.01888*, 2018.

Steve Berry, Oliver B Linton, and Ariel Pakes. Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654, 2004.

Steven Berry and Panle Jia. Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, 2(3):1–43, 2010.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

References II

Steven T Berry and Philip A Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014.

Kaicheng Chen and Harold D Chiang. Cross-fitting-free debiased machine learning with multiway dependence. *arXiv preprint arXiv:2602.11333*, 2026.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 02 2018.

Giovanni Compiani. Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics*, 13(2):545–591, 2022.

Giovanni Compiani, Ilya Morozov, and Stephan Seiler. Demand estimation with text and image data. *The RAND Journal of Economics*, 2026.

References III

Francesco Decarolis, Maria Polyakova, and Stephen P Ryan. Subsidy design in privately provided social insurance: Lessons from medicare part d. *Journal of Political Economy*, 128 (5):1712–1752, 2020.

Fabian Dunker, Stefan Hoderlein, and Hiroaki Kaido. Nonparametric identification of random coefficients in aggregate demand models for differentiated products. *The Econometrics Journal*, 26(2):279–306, 2023.

Joachim Freyberger. Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181, 2015.

Benjamin J Gillen, Sergio Montero, Hyungsik Roger Moon, and Matthew Shum. BLP-2LASSO for aggregate discrete choice models with rich covariates. *The Econometrics Journal*, 22(3): 262–281, 2019.

Paul LE Grieco, Charles Murry, and Ali Yurukoglu. The evolution of market power in the us automobile industry. *The Quarterly Journal of Economics*, 139(2):1201–1253, 2024.

References IV

Zhentong Lu, Xiaoxia Shi, and Jing Tao. Semi-nonparametric estimation of random coefficients logit model for aggregate demand. *Journal of Econometrics*, 235(2):2245–2265, 2023.

Masayuki Sawada and Kohei Kawaguchi. Estimating high-dimensional discrete choice model of differentiated products with random coefficients. *arXiv preprint arXiv:2004.08791*, 2020.

Ao Wang. Sieve BLP: A semi-nonparametric model of demand for differentiated products. *Journal of Econometrics*, 235(2):325–351, 2023.