# Detecting misspecification in the distribution of Random Coefficients in the aggregate demand model

Max Lesellier, Hippolyte Boucher, Gökçe Gökkoca

Toulouse School of Economics

August 23, 2022

# Introduction: Motivation

- Since [Berry, 1994] and [Berry et al., 1995], the random coefficient logit demand model (BLP) has become the workhorse model for demand estimation ( [Nevo, 2000], [Petrin, 2002], [Gentzkow and Shapiro, 2006],...)
    - Only market level data required/ Random Coefficients (RC) allow for preference heterogeneity

# Introduction: Motivation

- Since [Berry, 1994] and [Berry et al., 1995], the random coefficient logit demand model (BLP) has become the workhorse model for demand estimation ( [Nevo, 2000], [Petrin, 2002], [Gentzkow and Shapiro, 2006],...)
  - Only market level data required/ Random Coefficients (RC) allow for preference heterogeneity

- Researchers impose strong restrictions on the distribution of RC (eg: normal/degenerate).

# Introduction: Motivation

- Since [Berry, 1994] and [Berry et al., 1995], the random coefficient logit demand model (BLP) has become the workhorse model for demand estimation ( [Nevo, 2000], [Petrin, 2002], [Gentzkow and Shapiro, 2006],...)
  - Only market level data required/ Random Coefficients (RC) allow for preference heterogeneity

- Researchers impose strong restrictions on the distribution of RC (eg: normal/degenerate).
  - Parametric restrictions pose serious credibility issues:
    - Contradictory micro evidence [Vij and Krueger, 2017], [Caputo et al., 2018] ▸ Details
    - Validity of counterfactuals? ▸ Simulations Counterfactuals
    - Interpretation of the estimated parameters? [McCulloch and Neuhaus, 2011], [Hui et al., 2021]

# Introduction: Motivation

- Since [Berry, 1994] and [Berry et al., 1995], the random coefficient logit demand model (BLP) has become the workhorse model for demand estimation ( [Nevo, 2000], [Petrin, 2002], [Gentzkow and Shapiro, 2006],...)
  - Only market level data required/ Random Coefficients (RC) allow for preference heterogeneity
- Researchers impose strong restrictions on the distribution of RC (eg: normal/degenerate).
  - Parametric restrictions pose serious credibility issues:
    - Contradictory micro evidence [Vij and Krueger, 2017], [Caputo et al., 2018] ▸ Details
    - Validity of counterfactuals? ▸ Simulations Counterfactuals
    - Interpretation of the estimated parameters? [McCulloch and Neuhaus, 2011], [Hui et al., 2021]
  - Semi parametric approaches are rarely adapted to demand estimation( [Compiani, 2018], [Lu et al., 2019], [Wang, 2021])
    - Curse of dimensionality, stringent data requirements, arduous to implement ...

# Introduction: Motivation

- Since [Berry, 1994] and [Berry et al., 1995], the random coefficient logit demand model (BLP) has become the workhorse model for demand estimation ( [Nevo, 2000], [Petrin, 2002], [Gentzkow and Shapiro, 2006],...)
  - Only market level data required/ Random Coefficients (RC) allow for preference heterogeneity

- Researchers impose strong restrictions on the distribution of RC (eg: normal/degenerate).
  - Parametric restrictions pose serious credibility issues:
    - Contradictory micro evidence [Vij and Krueger, 2017], [Caputo et al., 2018] ▸ Details
    - Validity of counterfactuals? ▸ Simulations Counterfactuals
    - Interpretation of the estimated parameters? [McCulloch and Neuhaus, 2011], [Hui et al., 2021]
  - Semi parametric approaches are rarely adapted to demand estimation( [Compiani, 2018], [Lu et al., 2019], [Wang, 2021])
    - Curse of dimensionality, stringent data requirements, arduous to implement ...

**Our approach: create tools to detect a wrong distribution of RC and to sequentially correct for it**

# Introduction: Contribution

This paper has 2 main contributions:

# Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**

# Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**
2. We tailor these instruments to undertake 3 distinct missions

## Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**
2. We tailor these instruments to undertake 3 distinct missions
   - **Test**: we develop a specification test on the distribution of RC

# Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**
2. We tailor these instruments to undertake 3 distinct missions
   - **Test**: we develop a specification test on the distribution of RC
   - **Model selection:** help the econometrician decide which variables necessitate a RC

# Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**
2. We tailor these instruments to undertake 3 distinct missions
   - **Test**: we develop a specification test on the distribution of RC
   - **Model selection:** help the econometrician decide which variables necessitate a RC
   - **Estimation:** we show how these instruments can improve the estimation of the dist. of RC

# Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**
2. We tailor these instruments to undertake 3 distinct missions
   - **Test**: we develop a specification test on the distribution of RC
   - **Model selection:** help the econometrician decide which variables necessitate a RC
   - **Estimation:** we show how these instruments can improve the estimation of the dist. of RC
     - New set of BLP instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019])

# Introduction: Contribution

This paper has 2 main contributions:

1. **We construct new instruments, which are designed to detect misspecification in the distribution of RC**
2. We tailor these instruments to undertake 3 distinct missions
   - **Test**: we develop a specification test on the distribution of RC
   - **Model selection:** help the econometrician decide which variables necessitate a RC
   - **Estimation:** we show how these instruments can improve the estimation of the dist. of RC
     - New set of BLP instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019])
3. Empirical application to study the effects on Welfare and $C0_2$ emissions of different taxation schemes on cars

# Introduction: Contribution

This paper has 2 main contributions:

1 **We construct new instruments, which are designed to detect misspecification in the distribution of RC**

2 We tailor these instruments to undertake 3 distinct missions

- **Test**: we develop a specification test on the distribution of RC
- **Model selection:** help the econometrician decide which variables necessitate a RC
- **Estimation:** we show how these instruments can improve the estimation of the dist. of RC
  - New set of BLP instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019])

3 Empirical application to study the effects on Welfare and $C0_2$ emissions of different taxation schemes on cars

- This paper relates to the literatures on BLP (estimation, asymptotics, instruments), on specification tests (in structural models) ▸ More literature

# Outline

# Outline

# The model: the random utility

Consumer $i$ derives utility $u_{ijt}$ from good $j \in \{1, \ldots, J\}$ in market $t \in \{1, \ldots, T\}$

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi_{jt}}_{\delta_{jt}} + x'_{2jt}v_i + \varepsilon_{ijt}$$

- $x_{1jt}$ vector of product characteristics with no consumer heterogeneity, $\beta$ represents preferences for $x_{1jt}$
- $\xi_{jt}$ is unobserved product quality
- $\delta_{jt}$ is the mean utility for product $j$, common to all consumers
- $x_{2jt}$ vector of product characteristics with consumer heterogeneity
  - $v_i$ vector of random coefficients which follows the distribution $f$
- $\varepsilon_{ijt}$ an iid EV1 independent preference shock

# The model: the random utility

Consumer $i$ derives utility $u_{ijt}$ from good $j \in \{1, \dots, J\}$ in market $t \in \{1, \dots, T\}$

$$u_{ijt} = \underbrace{x_{1jt}'\beta + \xi_{jt}}_{\delta_{jt}} + x_{2jt}'v_i + \varepsilon_{ijt}$$

- $x_{1jt}$ vector of product characteristics with no consumer heterogeneity, $\beta$ represents preferences for $x_{1jt}$
- $\xi_{jt}$ is unobserved product quality
- $\delta_{jt}$ is the mean utility for product $j$, common to all consumers
- $x_{2jt}$ vector of product characteristics with consumer heterogeneity
  - $v_i$ vector of random coefficients which follows the distribution $f$
- $\varepsilon_{ijt}$ an iid EV1 independent preference shock

# The model: the random utility

Consumer $i$ derives utility $u_{ijt}$ from good $j \in \{1, \ldots, J\}$ in market $t \in \{1, \ldots, T\}$

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi_{jt}}_{\delta_{jt}} + x'_{2jt}v_i + \varepsilon_{ijt}$$

- $x_{1jt}$ vector of product characteristics with no consumer heterogeneity, $\beta$ represents preferences for $x_{1jt}$
- $\xi_{jt}$ is unobserved product quality
- $\delta_{jt}$ is the mean utility for product $j$, common to all consumers
- $x_{2jt}$ vector of product characteristics with consumer heterogeneity
    - $v_i$ vector of random coefficients which follows the distribution $f$
- $\varepsilon_{ijt}$ an iid EV1 independent preference shock

# The model: the random utility

Consumer $i$ derives utility $u_{ijt}$ from good $j \in \{1, \ldots, J\}$ in market $t \in \{1, \ldots, T\}$

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi_{jt}}_{\delta_{jt}} + x'_{2jt} v_i + \varepsilon_{ijt}$$

- $x_{1jt}$ vector of product characteristics with no consumer heterogeneity, $\beta$ represents preferences for $x_{1jt}$
- $\xi_{jt}$ is unobserved product quality
- $\delta_{jt}$ is the mean utility for product $j$, common to all consumers
- $x_{2jt}$ vector of product characteristics with consumer heterogeneity
  - $v_i$ vector of random coefficients which follows the distribution $f$
- $\varepsilon_{ijt}$ an iid EV1 independent preference shock

# The model: the random utility

Consumer $i$ derives utility $u_{ijt}$ from good $j \in \{1, \ldots, J\}$ in market $t \in \{1, \ldots, T\}$

$$u_{ijt} = \underbrace{x'_{1jt}\beta + \xi_{jt}}_{\delta_{jt}} + x'_{2jt} v_i + \varepsilon_{ijt}$$

- $x_{1jt}$ vector of product characteristics with no consumer heterogeneity, $\beta$ represents preferences for $x_{1jt}$
- $\xi_{jt}$ is unobserved product quality
- $\delta_{jt}$ is the mean utility for product $j$, common to all consumers
- $x_{2jt}$ vector of product characteristics with consumer heterogeneity
  - $v_i$ vector of random coefficients which follows the distribution $f$
- $\varepsilon_{ijt}$ an iid EV1 independent preference shock

# The model: Data Generating Process

- Consumer chooses the product which maximizes his/her utility:

$$
\begin{aligned}
s_{jt} &= \mathbb{P}(\text{good } j \text{ is chosen in market } t | x_t, \xi_t) \\
&= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x_{1jt}'\beta + \xi_{jt} + x_{2jt}'v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x_{1kt}'\beta + \xi_{kt} + x_{2kt}'v\right\}} f(v)dv
\end{aligned}
$$

# The model: Data Generating Process

- Consumer chooses the product which maximizes his/her utility:

$$\begin{aligned} s_{jt} &= \mathbb{P}(\text{good } j \text{ is chosen in market } t | x_t, \xi_t) \\ &= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x_{1jt}'\beta + \xi_{jt} + x_{2jt}'v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x_{1kt}'\beta + \xi_{kt} + x_{2kt}'v\right\}} f(v)dv \end{aligned}$$

- Market shares are assumed to be generated by the true distribution $f$ and the true parameter $\beta$

# The model: Data Generating Process

- Consumer chooses the product which maximizes his/her utility:

$$\begin{aligned} s_{jt} &= \mathbb{P}(\text{good } j \text{ is chosen in market } t | x_t, \xi_t) \\ &= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{1jt}\beta + \xi_{jt} + x'_{2jt}v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v)dv \end{aligned}$$

- Market shares are assumed to be generated by the true distribution $f$ and the true parameter $\beta$

- **Identifying condition** $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ where $z_{jt}$ instruments (exogenous characteristics, cost shifters...)

# The model: Data Generating Process

- Consumer chooses the product which maximizes his/her utility:

$$
\begin{aligned}
s_{jt} &= \mathbb{P}(\text{good } j \text{ is chosen in market } t | x_t, \xi_t) \\
&= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{1jt}\beta + \xi_{jt} + x'_{2jt}v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v)dv
\end{aligned}
$$

- Market shares are assumed to be generated by the true distribution $f$ and the true parameter $\beta$

- **Identifying condition** $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ where $z_{jt}$ instruments (exogenous characteristics, cost shifters...)

- Example of endogenous variable in $x_t$: price

# The model: Data Generating Process

- Consumer chooses the product which maximizes his/her utility:

$$\begin{aligned} s_{jt} &= \mathbb{P}(\text{good } j \text{ is chosen in market } t | x_t, \xi_t) \\ &= \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{x'_{1jt}\beta + \xi_{jt} + x'_{2jt}v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x'_{1kt}\beta + \xi_{kt} + x'_{2kt}v\right\}} f(v)dv \end{aligned}$$

- Market shares are assumed to be generated by the true distribution $f$ and the true parameter $\beta$

- **Identifying condition** $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$ where $z_{jt}$ instruments (exogenous characteristics, cost shifters...)

- Example of endogenous variable in $x_t$: price

- The econometrician observes $(s_t, x_{2t}, x_{1t}, z_t)_{t=1,\dots,T}$ and wants to estimate $\beta$ and $f$

# Outline

# Inverse demand function and structural error

- For any $\tilde{f}$, define the demand function $\rho(\cdot, x_{2t}, \tilde{f})$:

$$\begin{aligned} \rho(\cdot, x_{2t}, \tilde{f}) : \ & \mathbb{R}^J \to [0,1]^J \\ & \delta \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{\delta + x'_{2jt} v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_k + x'_{2kt} v\right\}} \tilde{f}(v) dv \end{aligned}$$

# Inverse demand function and structural error

- For any $\tilde{f}$, define the demand function $\rho(\cdot, x_{2t}, \tilde{f})$:

$$\begin{aligned}\rho(\cdot, x_{2t}, \tilde{f}):\ & \mathbb{R}^J \to [0,1]^J \\ & \delta \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{\delta + x_{2jt}' v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_k + x_{2kt}' v\right\}} \tilde{f}(v) dv\end{aligned}$$

- **Inverse demand**: [Berry, 1994] shows by applying Brouwer's fixed point that for any $(s_t, x_{2t})$ and for any $\tilde{f}$, there exists a unique $\delta^* \in \mathbb{R}^J$ such that:

$$s_t = \rho(\delta^*, x_{2t}, \tilde{f}) \iff \rho^{-1}(s_t, x_{2t}, \tilde{f}) = \delta^*$$

# Inverse demand function and structural error

- For any $\tilde{f}$, define the demand function $\rho(\cdot, x_{2t}, \tilde{f})$:

$$\rho(\cdot, x_{2t}, \tilde{f}) : \mathbb{R}^J \to [0,1]^J$$

$$\delta \mapsto \int_{\mathbb{R}^{K_2}} \frac{\exp\left\{\delta + x_{2jt}' v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_k + x_{2kt}' v\right\}} \tilde{f}(v) dv$$

- **Inverse demand**: [Berry, 1994] shows by applying Brouwer's fixed point that for any $(s_t, x_{2t})$ and for any $\tilde{f}$, there exists a unique $\delta^* \in \mathbb{R}^J$ such that:

$$s_t = \rho(\delta^*, x_{2t}, \tilde{f}) \iff \rho^{-1}(s_t, x_{2t}, \tilde{f}) = \delta^*$$

- The structural error generated by $(\tilde{f}, \tilde{\beta})$ is $\xi_{jt}(\tilde{f}, \tilde{\beta}) = \rho_j^{-1}(s_t, x_{2t}, \tilde{f}) - x_{1jt}' \tilde{\beta}$

- The structural error is recovered numerically via a contraction mapping Back

# Identification

- Exploiting [Wang, 2021], we derive mild but sufficient conditions for the non-parametric identification of $f$

  ▸ Identification proposition ▸ Details

# Identification

- Exploiting [Wang, 2021], we derive mild but sufficient conditions for the non-parametric identification of $f$

▸ Identification proposition ▸ Details

$$(\tilde{f}, \tilde{\beta}) = (f, \beta) \Longleftrightarrow \quad \mathbb{E}[\xi_{jt}(\tilde{f}, \tilde{\beta})|z_{jt}] = 0 \textit{ as}$$

# Identification

- Exploiting [Wang, 2021], we derive mild but sufficient conditions for the non-parametric identification of $f$
  ▸ Identification proposition ▸ Details

$$(\tilde{f}, \tilde{\beta}) = (f, \beta) \iff \quad \mathbb{E}[\xi_{jt}(\tilde{f}, \tilde{\beta}) | z_{jt}] = 0 \ as$$

- Strongest condition: completeness condition on the instruments

# Identification

- Exploiting [Wang, 2021], we derive mild but sufficient conditions for the non-parametric identification of $f$

▸ Identification proposition ▸ Details

$$(\tilde{f}, \tilde{\beta}) = (f, \beta) \iff \quad \mathbb{E}[\xi_{jt}(\tilde{f}, \tilde{\beta})|z_{jt}] = 0 \ as$$

- Strongest condition: completeness condition on the instruments
- This identification result gives us confidence that under rather weak conditions, we can detect a wrong distribution of RC.

# Detecting a wrong distribution of RC: the role of instruments

- **Objective**: construct instruments $h_D(z_{jt})$, which are "informative" about the distribution of RC

# Detecting a wrong distribution of RC: the role of instruments

- **Objective**: construct instruments $h_D(z_{jt})$, which are "informative" about the distribution of RC

- Consider a simple test $H_0 : f = f_0 \quad$ (assuming: $\beta_0 = \beta$)

# Detecting a wrong distribution of RC: the role of instruments

- **Objective**: construct instruments $h_D(z_{jt})$, which are "informative" about the distribution of RC

- Consider a simple test $H_0 : f = f_0 \quad$ (assuming: $\beta_0 = \beta$)

- An instrument $h_D(z_{jt})$ is informative about the distribution of RC if it is:

# Detecting a wrong distribution of RC: the role of instruments

- **Objective**: construct instruments $h_D(z_{jt})$, which are "informative" about the distribution of RC

- Consider a simple test $H_0 : f = f_0 \quad$ (assuming: $\beta_0 = \beta$)

- An instrument $h_D(z_{jt})$ is informative about the distribution of RC if it is:

  - **Valid**: under $H_0$, the instruments must not be correlated the structural error: $\mathbb{E}[\xi_{jt}(\beta_0, f_0)h_D(z_{jt})] = 0$

    - A sufficient condition: take $h_D(z_{jt})$ function of exogenous variables $z_{jt}$

# Detecting a wrong distribution of RC: the role of instruments

- **Objective**: construct instruments $h_D(z_{jt})$, which are "informative" about the distribution of RC

- Consider a simple test $H_0 : f = f_0 \quad$ (assuming: $\beta_0 = \beta$)

- An instrument $h_D(z_{jt})$ is informative about the distribution of RC if it is:
  - **Valid**: under $H_0$, the instruments must not be correlated the structural error: $\mathbb{E}[\xi_{jt}(\beta_0, f_0)h_D(z_{jt})] = 0$
    - A sufficient condition: take $h_D(z_{jt})$ function of exogenous variables $z_{jt}$
  - **Powerful**: under $H_1 : f = f_a \neq f_0$, the instruments must be strongly correlated to the structural error: $\mathbb{E}[\xi_{jt}(\beta_0, f_0)h_D(z_{jt})] \neq 0$

# Detecting a wrong distribution of RC: the role of instruments

- **Objective**: construct instruments $h_D(z_{jt})$, which are "informative" about the distribution of RC

- Consider a simple test $H_0 : f = f_0$ (assuming: $\beta_0 = \beta$)

- An instrument $h_D(z_{jt})$ is informative about the distribution of RC if it is:
  - **Valid**: under $H_0$, the instruments must not be correlated the structural error: $\mathbb{E}[\xi_{jt}(\beta_0, f_0)h_D(z_{jt})] = 0$
    - A sufficient condition: take $h_D(z_{jt})$ function of exogenous variables $z_{jt}$
  - **Powerful**: under $H_1 : f = f_a \neq f_0$, the instruments must be strongly correlated to the structural error: $\mathbb{E}[\xi_{jt}(\beta_0, f_0)h_D(z_{jt})] \neq 0$

- **Roadmap**:
  1. We derive an expression for the ideal instrument when $f_a$ known and infinite data
  2. We derive 2 feasible approximations of this instrument ($f_a$ unknown or unspecified)

# Most Powerful Instrument (MPI)

- We are under $H_1 : f = f_a$. What would be the ideal instrument to reject $H_0$?

# Most Powerful Instrument (MPI)

- We are under $H_1 : f = f_a$. What would be the ideal instrument to reject $H_0$?
- Def: the Most Powerful Instrument maximizes correlation with $\xi_{jt}(f_0, \beta_0)$ in the class of measurable functions of $z_{jt}$.

# Most Powerful Instrument (MPI)

- We are under $H_1 : f = f_a$. What would be the ideal instrument to reject $H_0$?
- Def: the Most Powerful Instrument maximizes correlation with $\xi_{jt}(f_0, \beta_0)$ in the class of measurable functions of $z_{jt}$.
- Decomposition of the structural error:

$$\xi_{jt}(f_0, \beta_0) = \underbrace{\xi_{jt}(f_a, \beta_a)}_{\text{true error}} + \underbrace{\Delta_j(s_t, x_{2t}, f_0, f_a)}_{\text{correction term due to misspecification}}$$

We show MPI writes: $$h^*_D(z_{jt}) = \alpha \mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \qquad \forall \alpha \neq 0$$

# Most Powerful Instrument (MPI)

- We are under $H_1 : f = f_a$. What would be the ideal instrument to reject $H_0$?
- Def: the Most Powerful Instrument maximizes correlation with $\xi_{jt}(f_0, \beta_0)$ in the class of measurable functions of $z_{jt}$.
- Decomposition of the structural error:

$$\xi_{jt}(f_0, \beta_0) = \underbrace{\xi_{jt}(f_a, \beta_a)}_{\text{true error}} + \underbrace{\Delta_j(s_t, x_{2t}, f_0, f_a)}_{\text{correction term due to misspecification}}$$

We show MPI writes: $$h_D^*(z_{jt}) = \alpha \mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \qquad \forall \alpha \neq 0$$

- MPI designed to capture exogenous variation in the correction term

# Most Powerful Instrument (MPI)

- We are under $H_1 : f = f_a$. What would be the ideal instrument to reject $H_0$?
- Def: the Most Powerful Instrument maximizes correlation with $\xi_{jt}(f_0, \beta_0)$ in the class of measurable functions of $z_{jt}$.
- Decomposition of the structural error:

$$\xi_{jt}(f_0, \beta_0) = \underbrace{\xi_{jt}(f_a, \beta_a)}_{\text{true error}} + \underbrace{\Delta_j(s_t, x_{2t}, f_0, f_a)}_{\text{correction term due to misspecification}}$$

We show MPI writes: $$h_D^*(z_{jt}) = \alpha \mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \qquad \forall \alpha \neq 0$$

- MPI designed to capture exogenous variation in the correction term
- **Difficulty**: MPI is alternative specific! In practice, we don't know the true alternative!

## "Global" Approximation of the MPI

By construction: $$\Delta_j(s_t, x_{2t}, f_0, f_a) = \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a)$$

## "Global" Approximation of the MPI

By construction: $$\Delta_j(s_t, x_{2t}, f_0, f_a) = \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a)$$

We show:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left(\frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^{J}\exp\left\{\delta_{kt}+x_{2jk}' v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^{J}\exp\left\{\delta_{jt}^0+x_{2jk}' v\right\}} f_0(v)dv}\right) \quad \text{with} \quad \delta_{jt}^0 = \rho_j^{-1}(s_t, x_{2t}, f_0)$$

- **Challenges remaining to construct instruments**:

# "Global" Approximation of the MPI

By construction: $\Delta_j(s_t, x_{2t}, f_0, f_a) = \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a)$

We show:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left( \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^{J} \exp\left\{\delta_{kt} + x_{2jk}' v\right\}} f_a(v) dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^{J} \exp\left\{\delta_{jt}^0 + x_{2jk}' v\right\}} f_0(v) dv} \right) \qquad \text{with} \quad \delta_{jt}^0 = \rho_j^{-1}(s_t, x_{2t}, f_0)$$

- **Challenges remaining to construct instruments**:
  - Some quantities are unknown to the econometrician: $f_a$ , $\delta_{jt}$

# "Global" Approximation of the MPI

By construction: $$\Delta_j(s_t, x_{2t}, f_0, f_a) = \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a)$$

We show:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log \left( \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^J \exp\left\{\delta_{kt} + x_{2jk}' v\right\}} f_a(v) dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x_{2jt}' v)}{1+\sum_{k=1}^J \exp\left\{\delta_{jt}^0 + x_{2jk}' v\right\}} f_0(v) dv} \right) \quad \text{with} \quad \delta_{jt}^0 = \rho_j^{-1}(s_t, x_{2t}, f_0)$$

- **Challenges remaining to construct instruments**:
  - Some quantities are unknown to the econometrician: $f_a$ , $\delta_{jt}$
    - Replace $\delta_{jt}$ by a known "close" substitute $\delta_{jt}^0$

# "Global" Approximation of the MPI

By construction: $$\Delta_j(s_t, x_{2t}, f_0, f_a) = \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a)$$

We show:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left(\frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\delta_{kt}+x'_{2jk}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{jt}+x'_{2jk}v\right\}} f_0(v)dv}\right) \quad \text{with } \delta^0_{jt} = \rho_j^{-1}(s_t, x_{2t}, f_0)$$

- **Challenges remaining to construct instruments**:
  - Some quantities are unknown to the econometrician: $f_a$ , $\delta_{jt}$
    - Replace $\delta_{jt}$ by a known "close" substitute $\delta^0_{jt}$
    - Approximate $f_a$ with a discrete distribution (next slide)

# "Global" Approximation of the MPI

By construction: $$\Delta_j(s_t, x_{2t}, f_0, f_a) = \rho_j^{-1}(s_t, x_{2t}, f_0) - \rho_j^{-1}(s_t, x_{2t}, f_a)$$

We show:

$$\Delta_j(s_t, x_{2t}, f_0, f_a) = \log\left( \frac{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\delta_{kt}+x'_{2jk}v\right\}} f_a(v)dv}{\int_{\mathbb{R}^{K_2}} \frac{\exp(x'_{2jt}v)}{1+\sum_{k=1}^{J}\exp\left\{\delta^0_{jt}+x'_{2jk}v\right\}} f_0(v)dv} \right) \qquad \text{with } \delta^0_{jt} = \rho_j^{-1}(s_t, x_{2t}, f_0)$$

- **Challenges remaining to construct instruments**:
  - Some quantities are unknown to the econometrician: $f_a$ , $\delta_{jt}$
    - Replace $\delta_{jt}$ by a known "close" substitute $\delta^0_{jt}$
    - Approximate $f_a$ with a discrete distribution (next slide)
  - Some variables are endogenous: $\delta^0_{jt}$
    - Replace $\delta^0_{jt}$ with $\hat{\delta}^0_{jt} = \widehat{\mathbb{E}[\delta^0_{jt}|z_{jt}]}$

## Global approximation of the MPI: interval instruments

We replace $f_a$ by a discrete distribution

$$\int_{\mathbb{R}} \frac{\exp\left\{x'_{2jt}v\right\}}{1+\sum_{k=1}^{J}\exp\left\{\hat{\delta}^0_{kt}+x'_{2kt}v\right\}} f_a(v)dv \approx \sum_{l=1}^{L} \omega_l \frac{exp(x'_{2jt}v_l)}{1+\sum_{k=1}^{J} exp(\hat{\delta}^0_{kt}+x'_{2kt}v_l)}$$

with $\{v_l\}_{l=1,\ldots,L}$ $L$ points chosen in the support of $f_a$, and $\omega_l$ the **unknown** weights associated with each point

# Global approximation of the MPI: interval instruments

We replace $f_a$ by a discrete distribution

$$\int_{\mathbb{R}} \frac{\exp\left\{x'_{2jt} v\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\hat{\delta}^0_{kt} + x'_{2kt} v\right\}} f_a(v) dv \approx \sum_{l=1}^{L} \omega_l \frac{exp(x'_{2jt} v_l)}{1 + \sum_{k=1}^{J} exp(\hat{\delta}^0_{kt} + x'_{2kt} v_l)}$$

with $\{v_l\}_{l=1,\dots,L}$ $L$ points chosen in the support of $f_a$, and $\omega_l$ the **unknown** weights associated with each point

We have the following approximation of the MPI

$$\mathbb{E}[\Delta_j(s_t, x_{2t}, f_0, f_a)|z_{jt}] \approx \log\left(\sum_{l=1}^{L} \omega_l \ \pi_{j,l}(z_{jt})\right)$$

We propose to use $\pi_{j,l}(z_{jt})$ as our new instruments (**interval instruments**).

local approximation

# Outline

## General idea behind the test

- In practice, the econometrician chooses a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ for the dist. of RC
  - Notation: $\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) \equiv \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$

## General idea behind the test

- In practice, the econometrician chooses a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ for the dist. of RC
  - Notation: $\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) \equiv \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$

- Estimation of a **pseudo-true value** $\theta_0 = (\beta_0, \lambda_0)$:

$$\theta_0 \equiv \theta(\mathcal{F}_0) = \underset{\tilde{\theta}}{Argmin}\ \mathbb{E}[\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) h_E(z_{jt})'] W \mathbb{E}[h_E(z_{jt}) \xi_{jt}(\mathcal{F}_0, \tilde{\theta})]$$

$h_E(z_{jt})$ instruments for estimation, $W$ weighting matrix ▸ Details estimation

## General idea behind the test

- In practice, the econometrician chooses a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ for the dist. of RC
  - Notation: $\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) \equiv \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$

- Estimation of a **pseudo-true value** $\theta_0 = (\beta_0, \lambda_0)$:

$$\theta_0 \equiv \theta(\mathcal{F}_0) = \underset{\tilde{\theta}}{Argmin}\ \mathbb{E}[\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) h_E(z_{jt})'] W \mathbb{E}[h_E(z_{jt}) \xi_{jt}(\mathcal{F}_0, \tilde{\theta})]$$

$h_E(z_{jt})$ instruments for estimation, $W$ weighting matrix ▸ Details estimation

- **We want to test:** $H_0 : f \in \mathcal{F}_0$. If pseudo true value unique:

$$H_0 : f \in \mathcal{F}_0 \implies \mathbb{E}[\xi_{jt}(\mathcal{F}_0, \theta_0)|z_{jt}] = 0 \ \ as \implies H_0' \quad \forall h_D(\cdot), \ \ \mathbb{E}[h_D(z_{jt}) \xi_{jt}(\mathcal{F}_0, \theta_0)] = 0$$

## General idea behind the test

- In practice, the econometrician chooses a parametric family $\mathcal{F}_0 = \{f_0(\cdot|\tilde{\lambda}) : \tilde{\lambda} \in \Lambda_0\}$ for the dist. of RC
  - Notation: $\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) \equiv \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})$

- Estimation of a **pseudo-true value** $\theta_0 = (\beta_0, \lambda_0)$:

$$\theta_0 \equiv \theta(\mathcal{F}_0) = \underset{\tilde{\theta}}{Argmin}\ \mathbb{E}[\xi_{jt}(\mathcal{F}_0, \tilde{\theta}) h_E(z_{jt})'] W \mathbb{E}[h_E(z_{jt}) \xi_{jt}(\mathcal{F}_0, \tilde{\theta})]$$

$h_E(z_{jt})$ instruments for estimation, $W$ weighting matrix Details estimation

- **We want to test:** $H_0 : f \in \mathcal{F}_0$. If pseudo true value unique:

$$H_0 : f \in \mathcal{F}_0 \implies \mathbb{E}[\xi_{jt}(\mathcal{F}_0, \theta_0)|z_{jt}] = 0 \;\; as \implies H_0' \quad \forall h_D(\cdot), \;\; \mathbb{E}[h_D(z_{jt}) \xi_{jt}(\mathcal{F}_0, \theta_0)] = 0$$

- We propose a moment based test for $H_0'$

# Test: Practical implementation

1. The researcher chooses a specification $\mathcal{F}_0$, instruments $h_E$ and estimates $\theta_0$

# Test: Practical implementation

1 The researcher chooses a specification $\mathcal{F}_0$, instruments $h_E$ and estimates $\theta_0$

2 The researcher chooses $h_D$ and constructs test statistic:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) = \frac{1}{T} \widehat{\xi}(\mathcal{F}_0, \hat{\theta})' h_D(z) \; \hat{\Sigma} \; h_D(z)' \widehat{\xi}(\mathcal{F}_0, \hat{\theta})$$

with some weights $\hat{\Sigma}$

# Test: Practical implementation

1 The researcher chooses a specification $\mathcal{F}_0$, instruments $h_E$ and estimates $\theta_0$

2 The researcher chooses $h_D$ and constructs test statistic:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) = \frac{1}{T} \widehat{\xi}(\mathcal{F}_0, \hat{\theta})' h_D(z) \, \hat{\Sigma} \, h_D(z)' \widehat{\xi}(\mathcal{F}_0, \hat{\theta})$$

with some weights $\hat{\Sigma}$

3 Decision rule for $H_0' : \mathbb{E}[h_D(z_{jt})\xi_{jt}(\mathcal{F}_0, \theta_0)] = 0$:

Under $H_0'$, $\quad S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z'\Sigma Z \quad$ with $Z \sim \mathcal{N}(0, \Omega_0)$ Formula

Reject $H_0'$ at level $\alpha$ if $S(h_D, \mathcal{F}_0, \hat{\theta}) > q_{1-\alpha}$ with $q_{1-\alpha}$ the $1 - \alpha$ quantile of $Z'\Sigma Z$

# Test: Practical implementation

1. The researcher chooses a specification $\mathcal{F}_0$, instruments $h_E$ and estimates $\theta_0$

2. The researcher chooses $h_D$ and constructs test statistic:

$$S(h_D, \mathcal{F}_0, \hat{\theta}) = \frac{1}{T} \widehat{\xi}(\mathcal{F}_0, \hat{\theta})' h_D(z) \; \hat{\Sigma} \; h_D(z)' \widehat{\xi}(\mathcal{F}_0, \hat{\theta})$$

   with some weights $\hat{\Sigma}$

3. Decision rule for $H_0' : \mathbb{E}[h_D(z_{jt}) \xi_{jt}(\mathcal{F}_0, \theta_0)] = 0$:

   Under $H_0'$, $\quad S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z' \Sigma Z \quad$ with $Z \sim \mathcal{N}(0, \Omega_0)$ ▸ Formula

   Reject $H_0'$ at level $\alpha$ if $S(h_D, \mathcal{F}_0, \hat{\theta}) > q_{1-\alpha}$ with $q_{1-\alpha}$ the $1-\alpha$ quantile of $Z' \Sigma Z$

- **Special case**: $h_D = h_E$ then the test is a Sargan-Hansen over id test

## Test: asymptotic properties

- We study asymptotics when $T$ goes to $+\infty$

# Test: asymptotic properties

- We study asymptotics when $T$ goes to $+\infty$
- **We prove the validity and consistency of our test under mild regularity conditions on the DGP and approx.**

▸ Theorem ▸ Assumptions

# Test: asymptotic properties

- We study asymptotics when $T$ goes to $+\infty$
- **We prove the validity and consistency of our test under mild regularity conditions on the DGP and approx.**

  ▸ Theorem ▸ Assumptions

- Consistent against alternatives of the form $H_1' : \mathbb{E}[h_D(z_{jt})\xi_{jt}(\mathcal{F}_0, \theta_0)] \neq 0 \quad$ (not $H_1 : f \notin \mathcal{F}_0$)
  - Choice of $h_D$ critical to detect misspecification
  - We suggest to choose our interval instruments, which are designed to detect mistakes in the dist of RC

# Test: asymptotic properties

- We study asymptotics when $T$ goes to $+\infty$
- **We prove the validity and consistency of our test under mild regularity conditions on the DGP and approx.**

  Theorem Assumptions

- Consistent against alternatives of the form $H_1' : \mathbb{E}[h_D(z_{jt})\xi_{jt}(\mathcal{F}_0, \theta_0)] \neq 0 \quad (\text{not } H_1 : f \notin \mathcal{F}_0)$
  - Choice of $h_D$ critical to detect misspecification
  - We suggest to choose our interval instruments, which are designed to detect mistakes in the dist of RC
- Main challenge to prove result: many approximations in estimation of $\theta_0$ (numerical inversion, numerical approximation of the integral...)

  numerical approximations

# Theoretical properties of the MPI for the specification test

Why is the MPI a good instrument for our test?

1 **Consistency:** with the MPI, our test is consistent against $H_1 : f \notin \mathcal{F}_0$ ▸ proof

# Theoretical properties of the MPI for the specification test

Why is the MPI a good instrument for our test?

1. **Consistency:** with the MPI, our test is consistent against $H_1 : f \notin \mathcal{F}_0$ ▸ proof

2. **Power in finite sample**

# Theoretical properties of the MPI for the specification test

Why is the MPI a good instrument for our test?

1. **Consistency:** with the MPI, our test is consistent against $H_1 : f \notin \mathcal{F}_0$ ▸ proof

2. **Power in finite sample**

   - **Power criteria:** Non local approach from [Bahadur, 1960]: choose the test which minimizes the level $\alpha$ required to attain a given power against a fixed alternative $f_a$.

# Theoretical properties of the MPI for the specification test

Why is the MPI a good instrument for our test?

1 **Consistency:** with the MPI, our test is consistent against $H_1 : f \notin \mathcal{F}_0$ proof

2 **Power in finite sample**

- **Power criteria:** Non local approach from [Bahadur, 1960]: choose the test which minimizes the level $\alpha$ required to attain a given power against a fixed alternative $f_a$.
- In practice, compute the asymptotic slopes using [Geweke, 1981]:

$$c_{h_D}(f_a) = plim \ \frac{1}{T} \underbrace{S(h_D, \mathcal{F}_0, \theta_0)}_{\text{test statistic}}$$

# Theoretical properties of the MPI for the specification test

Why is the MPI a good instrument for our test?

1. **Consistency:** with the MPI, our test is consistent against $H_1 : f \notin \mathcal{F}_0$ proof

2. **Power in finite sample**

- **Power criteria:** Non local approach from [Bahadur, 1960]: choose the test which minimizes the level $\alpha$ required to attain a given power against a fixed alternative $f_a$.
- In practice, compute the asymptotic slopes using [Geweke, 1981]:

$$c_{h_D}(f_a) = plim \ \frac{1}{T} \underbrace{S(h_D, \mathcal{F}_0, \theta_0)}_{\text{test statistic}}$$

- It can also be interpreted as measure of speed of divergence in terms of population moments: $Speed \approx T \times c_{h_D}(f_a)$

# Theoretical properties of the MPI for the specification test

Why is the MPI a good instrument for our test?

1. **Consistency:** with the MPI, our test is consistent against $H_1 : f \notin \mathcal{F}_0$ proof

2. **Power in finite sample**

   - **Power criteria:** Non local approach from [Bahadur, 1960]: choose the test which minimizes the level $\alpha$ required to attain a given power against a fixed alternative $f_a$.
   - In practice, compute the asymptotic slopes using [Geweke, 1981]:

$$c_{h_D}(f_a) = plim\ \frac{1}{T} \underbrace{S(h_D, \mathcal{F}_0, \theta_0)}_{\text{test statistic}}$$

   - It can also be interpreted as measure of speed of divergence in terms of population moments: $Speed \approx T \times c_{h_D}(f_a)$
   - We show that the slope is maximized (under homoskedasticity) by the MPI

# Monte Carlo simulations: Specification Test

- **Objective:** study the finite sample properties of our test and compare the performance of interval instruments with other instruments in the literature ▸ simulation design

# Monte Carlo simulations: Specification Test

- **Objective:** study the finite sample properties of our test and compare the performance of interval instruments with other instruments in the literature ▸ simulation design
- **Size of the test:** probability to reject $H_0 : f \in \mathcal{F}_0$ when $H_0$ is true.
  - We simulate data with $f$ normal and test normality of $f$
  - Our simulations show that empirical size converges to the nominal size ▸ size

# Monte Carlo simulations: Specification Test

- **Objective:** study the finite sample properties of our test and compare the performance of interval instruments with other instruments in the literature simulation design
- **Size of the test:** probability to reject $H_0 : f \in \mathcal{F}_0$ when $H_0$ is true.
  - We simulate data with $f$ normal and test normality of $f$
  - Our simulations show that empirical size converges to the nominal size size
- **Power of the test:** probability to reject $H_0 : f \in \mathcal{F}_0$ under $H_1 : f \notin \mathcal{F}_0$
  - We simulate data with $f$ not normal and test normality of $f$
  - Our simulations show interval instruments outperform the traditional instruments in term of power ("differentiation" and "optimal') power

# Outline

# Model selection

- **Motivation**: We want to know which variables display consumer heterogeneity and require a RC
  - without having to estimate model for each possible combination!!!

# Model selection

- **Motivation**: We want to know which variables display consumer heterogeneity and require a RC
  - without having to estimate model for each possible combination!!!

- **Intuition**: under misspecification, the structural error contains information about the true dist. of RC

# Model selection

- **Motivation**: We want to know which variables display consumer heterogeneity and require a RC
  - without having to estimate model for each possible combination!!!

- **Intuition**: under misspecification, the structural error contains information about the true dist. of RC

- **Pedagogical Example**: $x_{2jt} = (x^*_{2jt}, x^\dagger_{2jt})$. Only one characteristic displays consumer heterogeneity!

# Model selection

- **Motivation**: We want to know which variables display consumer heterogeneity and require a RC
  - without having to estimate model for each possible combination!!!

- **Intuition**: under misspecification, the structural error contains information about the true dist. of RC

- **Pedagogical Example**: $x_{2jt} = (x^*_{2jt}, x^\dagger_{2jt})$. Only one characteristic displays consumer heterogeneity!
  - Reference model: $H_0$ : "no consumer heterogeneity"

# Model selection

- **Motivation**: We want to know which variables display consumer heterogeneity and require a RC
  - without having to estimate model for each possible combination!!!

- **Intuition**: under misspecification, the structural error contains information about the true dist. of RC

- **Pedagogical Example**: $x_{2jt} = (x^*_{2jt}, x^\dagger_{2jt})$. Only one characteristic displays consumer heterogeneity!
  - Reference model: $H_0$ : "no consumer heterogeneity"
  - 2 competing alternatives: $H_1^*$ : " $x^*_{2jt}$ presents consumer heterogeneity" vs $H_1^\dagger$ : " $x^\dagger_{2jt}$ presents consumer heterogeneity"

# Model selection

- **Motivation**: We want to know which variables display consumer heterogeneity and require a RC
  - without having to estimate model for each possible combination!!!

- **Intuition**: under misspecification, the structural error contains information about the true dist. of RC

- **Pedagogical Example**: $x_{2jt} = (x^*_{2jt}, x^\dagger_{2jt})$. Only one characteristic displays consumer heterogeneity!
  - Reference model: $H_0$ : "no consumer heterogeneity"
  - 2 competing alternatives: $H_1^*$ : " $x^*_{2jt}$ presents consumer heterogeneity" vs $H_1^\dagger$ : " $x^\dagger_{2jt}$ presents consumer heterogeneity"
  - Reminder: $\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)$ the Structural Error under $H_0$ (estimated via simple 2SLS)

# Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = x'_{1jt}\gamma_1 + x^{\dagger}_{2jt}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x^*_{2t}, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] + u^*_{jt} \quad \text{with} \quad \mathbb{E}[u^*_{jt}|z_{jt}] = 0$$

# Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = x'_{1jt}\gamma_1 + x^{\dagger}_{2jt}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x^*_{2t}, f^*_0(\cdot|\lambda^*_0), f^*_a)|z_{jt}] + u^*_{jt} \quad \text{with} \quad \mathbb{E}[u^*_{jt}|z_{jt}] = 0$$

# Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0),\beta_0) = x'_{1jt}\gamma_1 + x^{\dagger}_{2jt}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x^*_{2t}, f^*_0(\cdot|\lambda^*_0), f^*_a)|z_{jt}] + u^*_{jt} \quad \text{with} \quad \mathbb{E}[u^*_{jt}|z_{jt}] = 0$$

  - Earlier, we showed $\mathbb{E}[\Delta_j(s_t, x^*_{2t}, f^*_0(\cdot|\lambda^*_0), f^*_a)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l \pi^*_{j,l}(z_{jt})$.

# Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = x'_{1jt}\gamma_1 + x^{\dagger}_{2jt}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x^*_{2t}, f^*_0(\cdot|\lambda^*_0), f^*_a)|z_{jt}] + u^*_{jt} \quad \text{with} \quad \mathbb{E}[u^*_{jt}|z_{jt}] = 0$$

  - Earlier, we showed $\mathbb{E}[\Delta_j(s_t, x^*_{2t}, f^*_0(\cdot|\lambda^*_0), f^*_a)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l \pi^*_{j,l}(z_{jt})$.

## Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = x_{1jt}'\gamma_1 + x_{2jt}^{\dagger}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x_{2t}^*, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] + u_{jt}^* \quad \text{with} \quad \mathbb{E}[u_{jt}^*|z_{jt}] = 0$$

  - Earlier, we showed $\mathbb{E}[\Delta_j(s_t, x_{2t}^*, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l \pi_{j,l}^*(z_{jt})$.

- Under $H_1^*$, a good linear regression model for $\xi_{jt}(f_0, \beta_0)$ is $M_1^*$ :

$$M_1^* : \qquad \xi_{jt}(\delta_{\lambda_0}, \beta_0) = x_{1jt}'\alpha_1 + x_{2jt}^{\dagger}\alpha_2 + \sum_{l=1}^{L} \omega_l \pi_{j,l}^*(z_{jt}) + u_{jt}^*$$

# Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = x'_{1jt}\gamma_1 + x^{\dagger}_{2jt}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x^*_{2t}, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] + u^*_{jt} \quad \text{with} \quad \mathbb{E}[u^*_{jt}|z_{jt}] = 0$$

  - Earlier, we showed $\mathbb{E}[\Delta_j(s_t, x^*_{2t}, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l \pi^*_{j,l}(z_{jt})$.

- Under $H_1^*$, a good linear regression model for $\xi_{jt}(f_0, \beta_0)$ is $M_1^*$ :

$$M_1^* : \qquad \xi_{jt}(\delta_{\lambda_0}, \beta_0) = x'_{1jt}\alpha_1 + x^{\dagger}_{2jt}\alpha_2 + \sum_{l=1}^{L} \omega_l \pi^*_{j,l}(z_{jt}) + u^*_{jt}$$

- Conversely, under $H_1^{\dagger}$ a good linear regression model writes:

$$M_1^{\dagger} : \qquad \xi_{jt}(\delta_{\lambda_0}, \beta_0) = x'_{1jt}\gamma_1 + x^*_{2jt}\gamma_2 + \sum_{l=1}^{L} \omega_l \pi^{\dagger}_{j,l}(z_{jt}) + u^{\dagger}_{jt}$$

# Selecting between 2 alternatives

- Under $H_1^*$, we can show:

$$\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) = x_{1jt}'\gamma_1 + x_{2jt}^{\dagger}\gamma_2 + \mathbb{E}[\Delta_j(s_t, x_{2t}^*, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] + u_{jt}^* \quad \text{with} \quad \mathbb{E}[u_{jt}^*|z_{jt}] = 0$$

  - Earlier, we showed $\mathbb{E}[\Delta_j(s_t, x_{2t}^*, f_0^*(\cdot|\lambda_0^*), f_a^*)|z_{jt}] \approx \sum_{l=1}^{L} \omega_l \pi_{j,l}^*(z_{jt})$.

- Under $H_1^*$, a good linear regression model for $\xi_{jt}(f_0, \beta_0)$ is $M_1^*$ :

$$M_1^* : \qquad \xi_{jt}(\delta_{\lambda_0}, \beta_0) = x_{1jt}'\alpha_1 + x_{2jt}^{\dagger}\alpha_2 + \sum_{l=1}^{L} \omega_l \pi_{j,l}^*(z_{jt}) + u_{jt}^*$$

- Conversely, under $H_1^{\dagger}$ a good linear regression model writes:

$$M_1^{\dagger} : \qquad \xi_{jt}(\delta_{\lambda_0}, \beta_0) = x_{1jt}'\gamma_1 + x_{2jt}^{*}\gamma_2 + \sum_{l=1}^{L} \omega_l \pi_{j,l}^{\dagger}(z_{jt}) + u_{jt}^{\dagger}$$

- Select the most relevant alternative by selecting the best fitting model ($R^2$, AIC,...)

# Sequential selection procedure

1. The econometrician estimates $\theta_0$ under a parametric restriction $H_0 : f \in \mathcal{F}_0$ (Initial stage $\mathcal{F}_0$ degenerate $\implies$ logit)

# Sequential selection procedure

1 The econometrician estimates $\theta_0$ under a parametric restriction $H_0 : f \in \mathcal{F}_0$ (Initial stage $\mathcal{F}_0$ degenerate $\implies$ logit)

2 The econometrician selects one alternative of the form $H_1^k : f_k \notin \mathcal{F}_{0,k}$ following procedure in previous slide

# Sequential selection procedure

1 The econometrician estimates $\theta_0$ under a parametric restriction $H_0 : f \in \mathcal{F}_0$ (Initial stage $\mathcal{F}_0$ degenerate $\implies$ logit)

2 The econometrician selects one alternative of the form $H_1^k : f_k \notin \mathcal{F}_{0,k}$ following procedure in previous slide

3 Test $H_0 : f \in \mathcal{F}_0$ against the chosen alternative $H_1^{\tilde{k}}$ by applying our specification test

# Sequential selection procedure

1. The econometrician estimates $\theta_0$ under a parametric restriction $H_0 : f \in \mathcal{F}_0$ (Initial stage $\mathcal{F}_0$ degenerate $\implies$ logit)
2. The econometrician selects one alternative of the form $H_1^k : f_k \notin \mathcal{F}_{0,k}$ following procedure in previous slide
3. Test $H_0 : f \in \mathcal{F}_0$ against the chosen alternative $H_1^{\tilde{k}}$ by applying our specification test

Repeat stages 1-3 until the test no longer rejects $H_0 : f \in \mathcal{F}_0$ or when the econometrician decides to stop (estimation becomes intractable...)

# Outline

# Conclusion

Misspecification in the distribution of RC can substantially affect counterfactuals in the BLP demand model

# Conclusion

Misspecification in the distribution of RC can substantially affect counterfactuals in the BLP demand model

- In this paper, we construct novel instruments which are designed to detect misspecification in the distribution of RC
- We develop a moment-based specification test for the distribution of the RC

# Conclusion

Misspecification in the distribution of RC can substantially affect counterfactuals in the BLP demand model

- In this paper, we construct novel instruments which are designed to detect misspecification in the distribution of RC
- We develop a moment-based specification test for the distribution of the RC
- We show these interval instruments can be effectively used to estimate the dist of RC

# Conclusion

Misspecification in the distribution of RC can substantially affect counterfactuals in the BLP demand model

- In this paper, we construct novel instruments which are designed to detect misspecification in the distribution of RC
- We develop a moment-based specification test for the distribution of the RC
- We show these interval instruments can be effectively used to estimate the dist of RC
- We provide a procedure to select the variables, which display consumer heterogeneity

# Conclusion

Misspecification in the distribution of RC can substantially affect counterfactuals in the BLP demand model

- In this paper, we construct novel instruments which are designed to detect misspecification in the distribution of RC
- We develop a moment-based specification test for the distribution of the RC
- We show these interval instruments can be effectively used to estimate the dist of RC
- We provide a procedure to select the variables, which display consumer heterogeneity
- We use these new instruments to compare the effects on pollution of different taxation schemes in the German car market

# Outline

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$
- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...

## Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$
- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...
    - Should the interval instruments be used to estimate the model?

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$

- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...
    - Should the interval instruments be used to estimate the model?

- **The traditional approach:** the literature proposes several approximations of the optimal instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019], [Conlon and Gortmaker, 2019],...)

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$
- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...
  - Should the interval instruments be used to estimate the model?
- **The traditional approach:** the literature proposes several approximations of the optimal instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019], [Conlon and Gortmaker, 2019],...)
  - We show declining performance of these instruments when distribution becomes more complex (eg: mixture)

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$
- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...
  - Should the interval instruments be used to estimate the model?
- **The traditional approach:** the literature proposes several approximations of the optimal instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019], [Conlon and Gortmaker, 2019],...)
  - We show declining performance of these instruments when distribution becomes more complex (eg: mixture)

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$
- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...
  - Should the interval instruments be used to estimate the model?
- **The traditional approach:** the literature proposes several approximations of the optimal instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019], [Conlon and Gortmaker, 2019],...)
  - We show declining performance of these instruments when distribution becomes more complex (eg: mixture)
- Approximating $\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta^*)}{\partial \theta}\middle| z_{jt}\right]$ can be challenging:

# Estimation

- Estimation: $f \in \mathcal{F}_0$ and we want to estimate true parameter $\theta^*$
- **Motivation**: parameters associated with RC are notoriously difficult to estimate : [Knittel and Metaxoglou, 2008]...
  - Should the interval instruments be used to estimate the model?
- **The traditional approach:** the literature proposes several approximations of the optimal instruments ( [Reynaert and Verboven, 2014], [Gandhi and Houde, 2019], [Conlon and Gortmaker, 2019],...)
  - We show declining performance of these instruments when distribution becomes more complex (eg: mixture)
- Approximating $\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta^*)}{\partial \theta}\middle| z_{jt}\right]$ can be challenging:
  - large dimension on $z_{jt}$
  - requires good first stage estimate of $\theta^*$

# MPI and Estimation

- Standard estimation can be interpreted as finding $\hat{\theta}$, which minimizes test statistic associated with $H_0 : \theta^* = \theta_0$

# MPI and Estimation

- Standard estimation can be interpreted as finding $\hat{\theta}$, which minimizes test statistic associated with $H_0 : \theta^* = \theta_0$
- We can show that the MPI maximizes the asymptotic slope of the GMM objective function when evaluated at parameter $\theta_0 \neq \theta^*$

# MPI and Estimation

- Standard estimation can be interpreted as finding $\hat{\theta}$, which minimizes test statistic associated with $H_0 : \theta^* = \theta_0$
- We can show that the MPI maximizes the asymptotic slope of the GMM objective function when evaluated at parameter $\theta_0 \neq \theta^*$
  - The optimal instrument is a local approximation of the MPI around $\theta^*$ detail

# MPI and Estimation

- Standard estimation can be interpreted as finding $\hat{\theta}$, which minimizes test statistic associated with $H_0 : \theta^* = \theta_0$

- We can show that the MPI maximizes the asymptotic slope of the GMM objective function when evaluated at parameter $\theta_0 \neq \theta^*$
  - The optimal instrument is a local approximation of the MPI around $\theta^*$ detail

- Challenge: the MPI and interval instruments are defined for a fixed candidate $\theta_0$ (whereas in estimation procedure: many candidates: $\theta_1$, $\theta_2$,...)

  - Slightly modify our interval instruments to make them suitable for estimation

## MPI and Estimation

- Standard estimation can be interpreted as finding $\hat{\theta}$, which minimizes test statistic associated with $H_0 : \theta^* = \theta_0$
- We can show that the MPI maximizes the asymptotic slope of the GMM objective function when evaluated at parameter $\theta_0 \neq \theta^*$
  - The optimal instrument is a local approximation of the MPI around $\theta^*$ detail
- Challenge: the MPI and interval instruments are defined for a fixed candidate $\theta_0$ (whereas in estimation procedure: many candidates: $\theta_1$, $\theta_2$,...)
  - Slightly modify our interval instruments to make them suitable for estimation
- Discussion of merits and weaknesses of taking interval vs approximations of the optimal instruments further discussion

# Monte Carlo simulations: Estimation

- **Objective:** Compare the finite sample performance of our interval instruments with other instruments in the literature
  ▸ simulation design

We simulate data with $f$ Gaussian and Gaussian mixture and estimate the parameters with different sets of instruments

- Our simulations show:
  - Similar performance between the three sets of instruments when we estimate a simple Gaussian as RC ▸ estimation gaussian
  - Interval instruments outperform the traditional instruments when $f$ is a Gaussian mixture ▸ estimation mixture

# Micro evidence

With individual data, there are several studies that highlight multi-modal preference distributions (following [Train, 2016] estimation procedure).

- [Caputo et al., 2018] uses data from choice experiments in the US and shows that willingness to pay for meat characteristics such as Certified US product or Guaranteed tender follow a bi-modal distribution.
- [Vij and Krueger, 2017] uses household travel survey data from San Francisco Bay Area, United States and show that value for time in-vehicle, and for walking, biking, waiting are not normally distributed and found to be asymmetric or bi-modal in the case of biking.
- Also high rates of rejection of the J test which is a specification test

Back

# Simulation evidence 1

## Principle

- Principle: We simulate the BLP model but estimate it by making the wrong assumption on the random coefficients we either assume that there is no random coefficient at all (logit), either that the random coefficient follows a normal.
- Focusing on 1 product and we compute the price elasticities and cross-price elasticies using the 2 estimators and compare them with the true price and true cross-price elasticities.

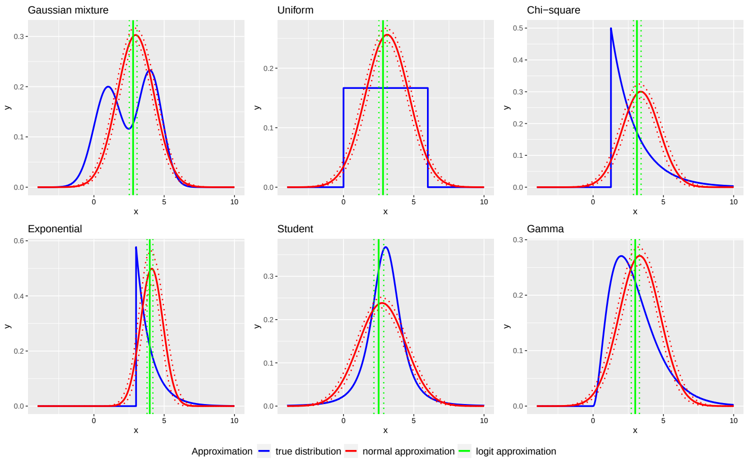
$\Rightarrow$ Most notably estimated cross-price elasticities are completely wrong therefore substitution parterns are completely wrong.

- DGP is the same as in other simulations DGP Back

# Simulation evidence 2

## Distribution of random coefficients

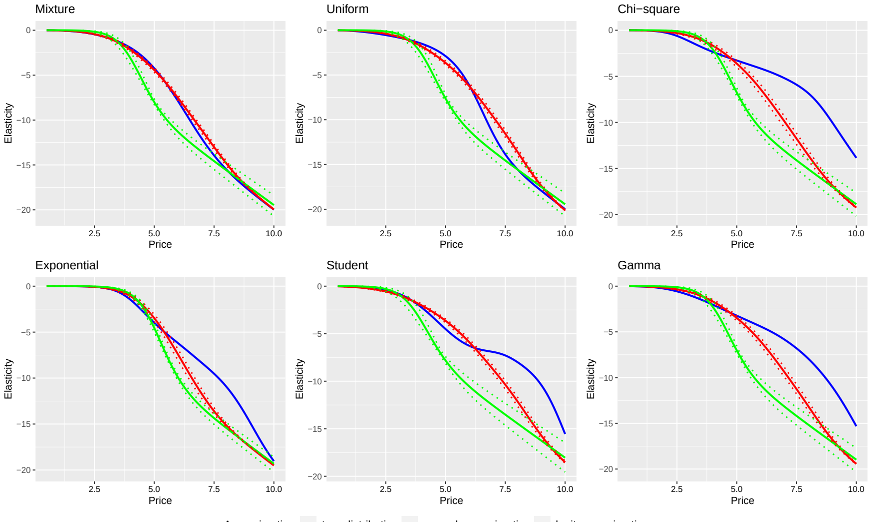
Figure 1: Approximation of Densities



# Simulation evidence 3

## Effect on price elasticities

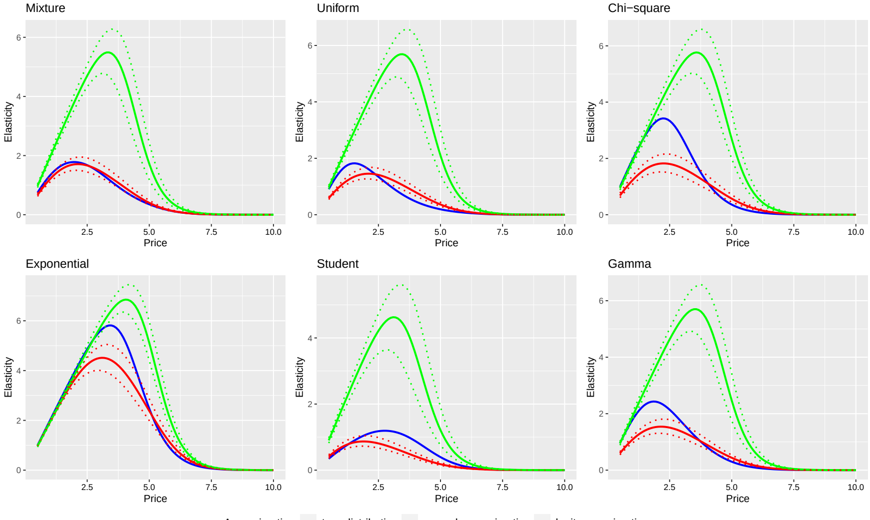
Figure 2: Price Elasticity



# Simulation evidence 4

## Effect on cross-price elasticities

Figure 3: Cross Price elasticity



## Related literature

- Non-parametric identification of random coefficients in BLP models: [Fox and Gandhi, 2016], [Fox et al., 2012], [Wang, 2020]
- Flexible estimation of random coefficient in BLP models: [Lu et al., 2019], [Compiani, 2018], [Fox et al., 2011]
- Practical implementation of BLP estimation: [Skrainka and Judd, 2011], [Dubé et al., 2012], [Reynaert and Verboven, 2014], [Lee and Seo, 2015], [Conlon and Gortmaker, 2019], [Gandhi and Houde, 2019]
- Non-normality of random coefficient: [Fosgerau and Hess, 2009], [Vij and Krueger, 2017], [Caputo et al., 2018]
- Misspecification: [White, 1982], [**?**], [McCulloch and Neuhaus, 2011] [Andrews and Shapiro, 2017], [Hui et al., 2021]

Back

# NFP algorithm

1. Choose starting values $\tilde{\theta} = (\tilde{\beta}, \tilde{\lambda})$
2. Derive starting values for the mean utilities $\delta_0 = X_1\beta$
3. Solve the contraction $H$ times for all $(j, t)$

$$\delta_{jt} = \delta_{jt} + log(s_{jt}) - log(\hat{\rho}_j(\delta_t, x_{2t}, f_0(\cdot|\tilde{\lambda})))$$

   with $\hat{\rho}_j(\delta_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ an approximation of $\rho_j(\delta_t, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ to obtain $\hat{\delta}(s, x_2, f_0(\cdot|\tilde{\lambda}))$
4. Back out the linear parameters and obtain an estimate of the structural error using 2SLS

$$\hat{\delta}(s, x_2, f_0(\cdot|\tilde{\lambda})) = x\tilde{\beta} + \xi_{jt}, \qquad \hat{\xi}(s, x_2, f_0(\cdot|\tilde{\lambda})) = \hat{\delta}(s, x_2, f_0(\cdot|\tilde{\lambda})) - x_1\hat{\beta}(\mathcal{F}_0, \tilde{\lambda})$$

$$\hat{\beta}(\mathcal{F}_0, \tilde{\lambda}) = \left(x_1' h_E(z)(h_E(z)' h_E(z))^{-1} h_E(z)' x_1\right)^{-1} x_1' h_E(z)(h_E(z)' h_E(z))^{-1} h_E(z)' \hat{\delta}(s, x_2, f_0(\cdot|\tilde{\lambda}))$$

5. Outer loop minimization problem with respect to $\tilde{\lambda}$

$$\underset{\tilde{\lambda}}{Argmin}\ \hat{\xi}' h_E(z) \hat{W} h_E(z)' \hat{\xi}$$

Back

# Assumptions for identification

1. Strict exogeneity: $E[\xi_{jt}|z_{jt}] = 0 \; a.s$
2. Completeness: for any measurable function $g(\cdot)$ such that $\mathbb{E}[g(s_t, x_t)] < \infty$, if $\mathbb{E}[g(s_t, x_t)|z_t] = 0 \; a.s$ , then $g(s_t, x_t) = 0 \; a.s$
3. $P(s_t, x_{2t}, x_{1t}, z_t)$ is observed by the econometrician and market shares $s_t$ are generated by the model
4. $x_t$ is such that $P(x_t' x_t$ is positive definite$) > 0$ for any $t$
5. There exists some $\bar{x}_t$ in $Supp(x_t)$ and an open set $\mathcal{D} \subset \mathbb{R}^J$ such that $\delta_t = \bar{x}_{1t}\beta_0 + \xi_t$ varies on $\mathcal{D}$
6. Let $\mathcal{X} = \{x_t \in Supp(x_t) \mid x_t' x_t \; is \; dp\}$. We assume that that $P(\mathcal{X} > 0)$
7. $v_i \perp (x_t, \xi_t, \varepsilon_{ijt})$

Back

# Non-parametric identification

- The identification result below implies that under fairly weak conditions, the data identifies the distribution of random coefficients nonparametrically.

## Proposition

*Under the assumptions in A, the distribution of random coefficients f and the homogeneous preference parameters $\beta$ are non-parametrically identified.*

$$(\tilde{f}, \tilde{\beta}) = (f, \beta) \Leftrightarrow \quad \forall j \qquad \mathbb{E}[\xi_{jt}(\tilde{f}, \tilde{\beta})|z_{jt}] = \mathbb{E}\left[\rho_j^{-1}(s_t, x_{2t}, \tilde{f}) - x'_{1jt}\tilde{\beta}\middle| z_{jt}\right] = 0 \; as$$

- This identification result gives us confidence that under weak conditions, it is possible for the econometrician to detect a wrongly specified distribution.

Back

# Most Powerful Instrument: "Local" Approximation

By exploiting the properties of the inverse demand function( $\mathcal{C}^{\infty}$ and bijective in $s_t$), we derive a first order expansion of $\Delta(s_t, x_{2t}, f_0, f)$ around $f_0$:

$$\Delta(s_t, x_{2t}, f_0, f) = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0)}{\partial \delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \left[\frac{\exp(\delta_t^0 + x_{2t} v)}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt}^0 + x_{2jk} v\right\}} - \rho_j(\delta_t^0, x_{2t}, f_0)\right] f(v) + \mathcal{R}_0$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0)$ and $\mathcal{R}_0 = o\left(\int_{\mathbb{R}^{K_2}} |f(v) - f_0(v)| dv\right)$.

# Most Powerful Instrument: "Local" Approximation

By exploiting the properties of the inverse demand function( $\mathcal{C}^\infty$ and bijective in $s_t$), we derive a first order expansion of $\Delta(s_t, x_{2t}, f_0, f)$ around $f_0$:

$$\Delta(s_t, x_{2t}, f_0, f) = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0)}{\partial \delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \left[\frac{\exp(\delta_t^0 + x_{2t} v)}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt}^0 + x_{2jk} v\right\}} - \rho_j(\delta_t^0, x_{2t}, f_0)\right] f(v) + \mathcal{R}_0$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0)$ and $\mathcal{R}_0 = o\left(\int_{\mathbb{R}^{K_2}} |f(v) - f_0(v)| dv\right)$.

- Same challenges: $f$ unknown , $\delta_{jt}^0$ endogenous

# Most Powerful Instrument: "Local" Approximation

By exploiting the properties of the inverse demand function( $\mathcal{C}^\infty$ and bijective in $s_t$), we derive a first order expansion of $\Delta(s_t, x_{2t}, f_0, f)$ around $f_0$:

$$\Delta(s_t, x_{2t}, f_0, f) = \left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0)}{\partial \delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \left[\frac{\exp(\delta_t^0 + x_{2t} v)}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt}^0 + x_{2jk} v\right\}} - \rho_j(\delta_t^0, x_{2t}, f_0)\right] f(v) + \mathcal{R}_0$$

with $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0)$ and $\mathcal{R}_0 = o\left(\int_{\mathbb{R}^{K_2}} |f(v) - f_0(v)| dv\right)$.

- Same challenges: $f$ unknown , $\delta_{jt}^0$ endogenous
  - Same solutions as previously
- Local as the approximation is accurate when $f$ close to $f_0$ in the $L_1$ norm

global approximation

# Details "Local" Approximation of the MPI

We obtain this result by observing that for any density $f_0$, we can construct artificial market shares $s_t^0$ such that $\rho^{-1}(s_t, x_{2t}, f) = \rho^{-1}(s_t^0, x_{2t}, f_0)$ and then we can take a Taylor expansion of $\rho^{-1}(s_t^0, x_{2t}, f_0)$ around $s_t$.

# Remarks on the test

- Rejecting $H_0' : E(h_D(z_{jt})\xi_{jt}(f,\theta)) = 0$ implies rejecting $H_0$, ie $(f,\theta) \neq (f_0,\theta_0)$. But not rejecting $H_0'$ does not imply $H_0$, ie it does not imply $(f,\theta) = (f_0,\theta_0)$
- In practice the test may lose power in some cases
- Other tests can be considered (Score, ICM)
- Other types of misspecification (missing variables, heteroskedasticity, some nonlinearities in the indirect utility) do not generate correlation between $z_{jt}$ and $\xi_{jt}(f,\theta)$, random logit models are very general [McFadden and Train, 2000]
- IVs are exogenous by construction (BLP instruments) or assumption (cost shifters)

$\Rightarrow$ Idea: Find a better $h_D$ to maximize power, then we can determine when to increase flexibility of $\mathcal{F}_0$

Back

# Details test implementation

- Under the null $H_0 : f \in \mathcal{F}_0$, under assumptions (B)-(E) and for any $\hat{\Sigma}$ such that *plim* $\hat{\Sigma} = \Sigma$,

$$S(h_D, \mathcal{F}_0, \hat{\theta}) \xrightarrow{d} Z'\Sigma Z, \qquad Z \sim \mathcal{N}(0, \Omega_0)$$

where

$$\Omega_0 = \begin{pmatrix} I_{|h_D|_0} & G \end{pmatrix} \begin{pmatrix} \Omega(\mathcal{F}_0, h_D) & \Omega(\mathcal{F}_0, h_D, h_E) \\ \Omega(\mathcal{F}_0, h_D, h_E)' & \Omega(\mathcal{F}_0, h_E) \end{pmatrix} \begin{pmatrix} I_{|h_D|_0} \\ G' \end{pmatrix}$$

with

$$\Omega(\mathcal{F}_0, h_D, h_E) = cov\left( \sum_j \xi_{jt}(f(.|\lambda_0), \beta_0) h_D(z_{jt}), \sum_j \xi_{jt}(f(.|\lambda_0), \beta_0) h_E(z_{jt}) \right)$$

$$G = -\Gamma(\mathcal{F}_0, \theta_0, h_D) \left[ \Gamma(\mathcal{F}_0, \theta_0, h_E)' W \Gamma(\mathcal{F}_0, \theta_0, h_E) \right]^{-1} \Gamma(\mathcal{F}_0, \theta_0, h_E)' W$$

$$\Gamma(\mathcal{F}_0, \theta_0, h) = \mathbb{E}\left[ \sum_j h(z_{jt}) \frac{\partial \xi_{jt}(f_0(.|\lambda_0), \beta_0)}{\partial \theta'} \right]$$

Back

# Validity and consistency theorems

## Theorem

*Let* $\hat{\theta} = \hat{\theta}(\mathcal{F}_0, \hat{W}, h_E)$ *be the BLP estimator associated with distributional assumption* $\mathcal{F}_0$*, weighting matrix* $\hat{W}$*, estimating instruments* $h_E$*. Under assumptions (B)-(E)*

- *Under* $H_0 : f \in \mathcal{F}_0$

$$\mathbb{P}(S(h_D, \mathcal{F}_0, \hat{\theta}) > q_{1-\alpha}) \to \alpha$$

*where* $q_{1-\alpha}$ *is the* $1-\alpha$ *quantile of* $Z'\Sigma Z$ ▸ Formula

- *Under* $H_1' : \mathbb{E}\left[\sum_j h_D(z_{jt})\xi_{jt}(f_0(.|\lambda_0), \beta_0)\right] \neq 0$

$$\forall q \in \mathbb{R}^+ \quad \mathbb{P}(S(h_D, \mathcal{F}_0, \hat{\theta}) > q) \to 1$$

▸ Back

# Assumptions for validity and consistency 1

- First assumption is regular and ensures bounded 2nd moments of $(z_{jt}, x_{jt}, \xi_{jt})$
- Second assumption ensures estimation is possible assuming $f \in \mathcal{F}_0$

## Assumption (A)

*(i)* $(z_t, x_t, s_t)_{t=1}^{T}$ *are iid across markets such that the probability model holds at* $(f, \theta)$

*(ii) Exogeneity:* $\forall j \;\; \mathbb{E}[\rho_j^{-1}(s_t, x_{2t}, f) - x_{1jt}'\beta | z_{jt}] = 0$ *as*

*(iii) Finite moment conditions:* $x_{2t}$ *has bounded support and* $x_{1t}$ *has finite 4th moments*

## Assumption (C)

$\mathcal{F}_0$ *is such that*

*(i)* $\lambda_0$ *belongs to the interior of* $\Lambda_0$ *with* $\Lambda_0$ *compact*

*(ii)* $\tilde{\lambda} \mapsto \rho(\delta, x_{2t}, f_0(\cdot|\tilde{\lambda}))$ *is well defined and continuously differentiable on* $\Lambda_0$

*(iii)* $\forall(\lambda, \lambda')$ *such that* $\lambda \neq \lambda'$, $\exists v^* \in Supp(\mathcal{F}_0)$ *such that* $f_0(v^*|\lambda) \neq f_0(v^*|\lambda')$

# Assumptions for validity and consistency 2

- Third assumption ensures proper identification and estimation of $\theta_0$ and allows for inference

## Assumption (D)

*Given $\mathcal{F}_0$ which satisfy Assumption (C) and for some weighting matrix $W$ and $\Sigma$*
*(i) Finite IV moments: $h_E(z_{jt})$ and $h_D(z_{jt})$ are not perfectly colinear and have finite 4th moments*
*(ii) Local identification: $\Gamma(\mathcal{F}_0, \theta_0, h_E) = \mathbb{E}\left[\sum_j h_E(z_{jt}) \frac{\partial \xi_{jt}(f_0(.|\lambda_0), \beta_0)}{\partial \theta'}\right]$ and $\Gamma(\mathcal{F}_0, \theta_0, h_D)$ are full rank (ie of rank $|\theta_0|$)*
*(iii) Global identification of $\theta_0$: $\forall \tilde{\theta} \neq \theta_0$:*

$$\mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta}) h_E(z_{jt})'\right] W \mathbb{E}\left[\sum_j h_E(z_{jt}) \xi_{jt}(f_0(\cdot|\tilde{\lambda}), \tilde{\beta})\right] > \mathbb{E}\left[\sum_j \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0) h_E(z_{jt})'\right] W \mathbb{E}\left[\sum_j h_E(z_{jt}) \xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)\right]$$

*(iv) $W$ and $\Sigma$ are symmetric positive definite and $\hat{W} \xrightarrow{P} W$, $\hat{\Sigma} \xrightarrow{P} \Sigma$*
*(v) $\hat{\theta}$ minimizes the BLP objective and satisfies the FOC of the minimization problem:*

$$\frac{\partial \widehat{\xi}(f(.|\hat{\lambda}), \hat{\beta})}{\partial \theta}' h_E(z) \hat{W} \widehat{\xi}(f(.|\hat{\lambda}), \hat{\beta})' h_E(z) = 0$$

Back

# Numerical error assumption

Major difficulty is numerical approximations. 3 types of numerical approximations:

1. Integral in the demand has to be approximated
2. Fixed point: $s = \rho(\delta, x_2, f_0(\dot{|}\tilde{\lambda}))$ is never fully satisfied in practice
3. Observed market shares $\hat{s}$ are empirical probability masses, in practice there is a finite number of individuals in each market Back

# Numerical error assumption

Major difficulty is numerical approximations. 3 types of numerical approximations:

1. Integral in the demand has to be approximated
2. Fixed point: $s = \rho(\delta, x_2, f_0(\cdot|\tilde{\lambda}))$ is never fully satisfied in practice
3. Observed market shares $\hat{s}$ are empirical probability masses, in practice there is a finite number of individuals in each market

Back

## Assumption (E)

*Let $R$ be the number draws to compute $\rho$, $n_t$ the number of individuals in market $t$, $H$ the stopping time of NFP and $\varepsilon \in (0;1)$ the contraction constant in NFP*

$$\frac{T}{R} \underset{T\to+\infty}{\longrightarrow} 0, \quad \forall t \ \frac{T}{n_t} \underset{T\to+\infty}{\longrightarrow} 0, \quad \sqrt{T}\varepsilon^H \underset{T\to+\infty}{\longrightarrow} 0$$

# Proof of consistency

$$
\begin{aligned}
H_1: \ f \notin \mathcal{F}_0 \implies & \quad \mathbb{E}[\xi_{jt}(f_0(\cdot|\lambda_0), \beta_0)|z_{jt}] \neq 0 \ \ a.s \\
\implies & \quad \mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}]^2 > 0 \ a.s \\
\implies & \quad \mathbb{E}\left[\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)|z_{jt}]^2\right] > 0 \\
\implies & \quad \mathbb{E}\left[\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0)|z_{jt}]|z_{jt}]\right] > 0 \\
\implies & \quad \mathbb{E}\left[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\mathbb{E}[\xi_{jt}(f_0(.|\lambda_0)|z_{jt}]\right] > 0 \\
\implies & \ \forall\alpha \neq 0 \ \ H_1': \ \mathbb{E}\Big[\xi_{jt}(f_0(.|\lambda_0), \beta_0)\underbrace{\alpha\mathbb{E}[\Delta_{0,a}^{\xi_{jt}}|z_{jt}]}_{h_D^*(z_{jt})}\Big] > 0
\end{aligned}
$$

▸ back

# Simulations: set-up

- Setting close to [Dubé et al., 2012] and [Reynaert and Verboven, 2014] ▸ Implementation

  Indirect utility is given by

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt}, \quad \xi_{jt} \overset{iid}{\sim} \mathcal{N}(0,1), \quad \varepsilon_{ijt} \overset{iid}{\sim} EV(1)$$

  - $T \in \{50, 100, 200\}$ , $J = 12$
  - $x_a$, $x_b$, $x_c$ are normal and correlated
  - Price is endogenous $p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt}$

## Simulations: set-up

- Setting close to [Dubé et al., 2012] and [Reynaert and Verboven, 2014] ▸ Implementation

  Indirect utility is given by

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt}, \quad \xi_{jt} \overset{iid}{\sim} \mathcal{N}(0,1), \quad \varepsilon_{ijt} \overset{iid}{\sim} EV(1)$$

  - $T \in \{50, 100, 200\}$ , $J = 12$
  - $x_a$, $x_b$, $x_c$ are normal and correlated
  - Price is endogenous $p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt}$

- Estimation is always done assuming normality, ie $\mathcal{F}_0 = \mathcal{N}(\mu, \sigma^2)$
- $v_i \sim f$ and $f$ varies between a normal (size), mixture of normals (power), etc...

# Simulations: set-up

- Setting close to [Dubé et al., 2012] and [Reynaert and Verboven, 2014] ▸ Implementation

  Indirect utility is given by

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt}, \quad \xi_{jt} \overset{iid}{\sim} \mathcal{N}(0,1), \quad \varepsilon_{ijt} \overset{iid}{\sim} EV(1)$$

  - $T \in \{50, 100, 200\}$ , $J = 12$
  - $x_a$, $x_b$, $x_c$ are normal and correlated
  - Price is endogenous $p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt}$

- Estimation is always done assuming normality, ie $\mathcal{F}_0 = \mathcal{N}(\mu, \sigma^2)$
- $v_i \sim f$ and $f$ varies between a normal (size), mixture of normals (power), etc...
- We consider different sets instruments:
  - Differentiation instruments [Gandhi and Houde, 2019]; "Optimal instrument " [Reynaert and Verboven, 2014]; Interval instruments

▸ back

## Simulations: empirical size

- Size = probability to reject the null when the null is true. We work under the null $f$ is normal and check that empirical size equal nominal size.

- J test(1) = J test with differentiation IVs, J test(2) = J test with optimal IVs

  I test(1) test with interval IVs and differentiation IVs for estimation

  I test(2) test with interval IVs and optimal IVs for estimation ▸ Details ▸ back

# Simulations: empirical size

- Size = probability to reject the null when the null is true. We work under the null $f$ is normal and check that empirical size equal nominal size.

- J test(1) = J test with differentiation IVs, J test(2) = J test with optimal IVs

  I test(1) test with interval IVs and differentiation IVs for estimation

  I test(2) test with interval IVs and optimal IVs for estimation Details back

| **Number of markets** | T=50 | | | | T=100 | | | | T=200 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test type** | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) |
| $v_i \sim \mathcal{N}(-1, 0.5^2)$ | 0.201 | 0.102 | 0.128 | 0.07 | 0.098 | 0.067 | 0.085 | 0.039 | 0.067 | 0.051 | 0.073 | 0.044 |
| $v_i \sim \mathcal{N}(0, 0.75^2)$ | 0.204 | 0.103 | 0.132 | 0.087 | 0.105 | 0.066 | 0.085 | 0.044 | 0.066 | 0.052 | 0.072 | 0.042 |
| $v_i \sim \mathcal{N}(1, 1^2)$ | 0.199 | 0.101 | 0.134 | 0.076 | 0.106 | 0.064 | 0.089 | 0.046 | 0.072 | 0.051 | 0.074 | 0.036 |
| $v_i \sim \mathcal{N}(2, 2^2)$ | 0.199 | 0.11 | 0.138 | 0.084 | 0.107 | 0.07 | 0.093 | 0.056 | 0.069 | 0.051 | 0.078 | 0.047 |
| $v_i \sim \mathcal{N}(3, 3^2)$ | 0.191 | 0.116 | 0.129 | 0.091 | 0.101 | 0.074 | 0.09 | 0.059 | 0.077 | 0.073 | 0.076 | 0.056 |

Figure 4: Empirical Size for Nominal Size 5%

## Simulations: empirical size

- Size = probability to reject the null when the null is true. We work under the null $f$ is normal and check that empirical size equals nominal size.

- J test(1) = J test with differentiation IVs, J test(2) = J test with optimal IVs

  I test(1) test with interval IVs and differentiation IVs for estimation

  I test(2) test with interval IVs and optimal IVs for estimation  Details  back

| **Number of markets** | T=50 | | | | T=100 | | | | T=200 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test type** | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) |
| $v_i \sim \mathcal{N}(-1, 0.5^2)$ | 0.201 | **0.102** | 0.128 | **0.07** | 0.098 | **0.067** | 0.085 | **0.039** | 0.067 | **0.051** | 0.073 | **0.044** |
| $v_i \sim \mathcal{N}(0, 0.75^2)$ | 0.204 | **0.103** | 0.132 | **0.087** | 0.105 | **0.066** | 0.085 | **0.044** | 0.066 | **0.052** | 0.072 | **0.042** |
| $v_i \sim \mathcal{N}(1, 1^2)$ | 0.199 | **0.101** | 0.134 | **0.076** | 0.106 | **0.064** | 0.089 | **0.046** | 0.072 | **0.051** | 0.074 | **0.036** |
| $v_i \sim \mathcal{N}(2, 2^2)$ | 0.199 | **0.11** | 0.138 | **0.084** | 0.107 | **0.07** | 0.093 | **0.056** | 0.069 | **0.051** | 0.078 | **0.047** |
| $v_i \sim \mathcal{N}(3, 3^2)$ | 0.191 | **0.116** | 0.129 | **0.091** | 0.101 | **0.074** | 0.09 | **0.059** | 0.077 | **0.073** | 0.076 | **0.056** |

Figure 5: Empirical Size for Nominal Size 5%

# Simulations: Power against mixture of normals

- Power = probability to reject the null when the null is not true
- True distribution $f$ is a mixture of normals ▸ back

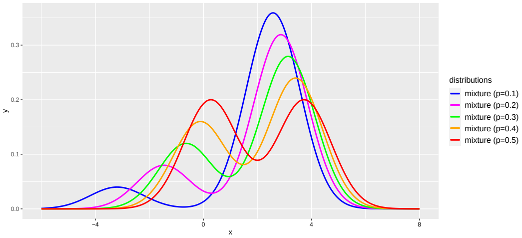



Figure 6: Densities of the True Distributions

# Simulations: Power against mixture of normals

- Power = probability to reject the null when the null is not true
- True distribution $f$ is a mixture of normals ▸ back

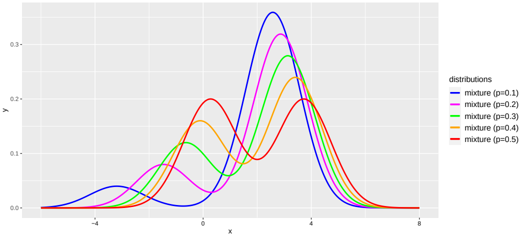



Figure 6: Densities of the True Distributions

| **Number of markets** | T=50 | | | | T=100 | | | | T=200 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test type** | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) |
| Mixture 1 | 0.257 | 0.997 | 0.726 | 0.993 | 0.185 | 1 | 0.964 | 0.999 | 0.225 | 1 | 1 | 1 |
| Mixture 2 | 0.279 | 1 | 0.589 | 0.999 | 0.221 | 1 | 0.919 | 0.999 | 0.277 | 1 | 0.999 | 1 |
| Mixture 3 | 0.312 | 0.996 | 0.397 | 0.993 | 0.251 | 1 | 0.704 | 1 | 0.326 | 1 | 0.981 | 1 |
| Mixture 4 | 0.338 | 0.984 | 0.236 | 0.973 | 0.289 | 1 | 0.375 | 0.997 | 0.404 | 1 | 0.684 | 1 |
| Mixture 5 | 0.347 | 0.925 | 0.142 | 0.905 | 0.326 | 0.997 | 0.111 | 1 | 0.458 | 1 | 0.162 | 1 |

Figure 7: Empirical Power, Gaussian Mixture Alternatives

# Simulations: Power against gamma distribution

- Power = probability to reject the null when the null is not true
- True distribution $f$ are gamma dist.

back

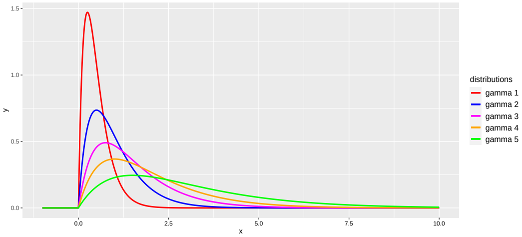



Figure 8: Densities of the True Distributions

# Simulations: Power against gamma distribution

- Power = probability to reject the null when the null is not true
- True distribution $f$ are gamma dist.

back

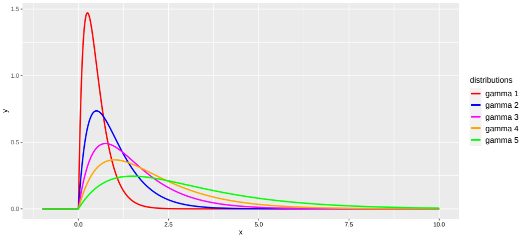



Figure 8: Densities of the True Distributions

| Number of markets | T=50 | | | | T=100 | | | | T=200 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test type** | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) | J test(1) | I test(1) | J test(2) | I test(2) |
| Gamma 1 | 0.194 | 0.12 | 0.133 | 0.088 | 0.12 | 0.092 | 0.086 | 0.082 | 0.101 | 0.15 | 0.069 | 0.132 |
| Gamma 2 | 0.428 | 0.752 | 0.131 | 0.737 | 0.495 | 0.965 | 0.092 | 0.963 | 0.798 | 1 | 0.088 | 1 |
| Gamma 3 | 0.489 | 0.958 | 0.155 | 0.964 | 0.606 | 1 | 0.131 | 1 | 0.883 | 1 | 0.176 | 1 |
| Gamma 4 | 0.449 | 0.996 | 0.217 | 0.992 | 0.551 | 1 | 0.259 | 1 | 0.801 | 1 | 0.437 | 1 |
| Gamma 5 | 0.415 | 1 | 0.36 | 0.997 | 0.468 | 1 | 0.55 | 0.999 | 0.705 | 1 | 0.872 | 1 |

Figure 9: Empirical Power, Gaussian Mixture Alternatives

- $f \in \mathcal{F}_0$, the MPI has the form: $h_D^*(z_{jt}) = \mathbb{E}\left[\Delta_{\theta_0,\theta^*}^{\xi_{jt}} | z_{jt}\right]$ with $\Delta_{\theta_0,\theta^*}^{\xi_{jt}} = \xi_{jt}(\theta_0) - \xi_{jt}(\theta^*)$.

- By taking a Taylor expansion of $\xi_{jt}(\theta_0)$ around $\theta^*$, we obtain:

$$\Delta_{\theta_0,\theta^*}^{\xi_{jt}} = \left[\frac{\partial \xi_{jt}(\theta^*)}{\partial \lambda}(\lambda_0 - \lambda^*) + x_{1jt}'(\beta^* - \beta_0)\right] + o(||\theta_0 - \theta^*||_2)$$

- $\theta_0$ is in a neighborhood of $\theta^*$, the MPI $h_D^*$ is a linear combination of the optimal instruments $h_E^*$.

$$h_D^*(z_{jt}) = \mathbb{E}\left[\Delta_{\theta_0,\theta^*}^{\xi_{jt}} | z_{jt}\right] \approx \underbrace{\mathbb{E}\left[\frac{\partial \xi_{jt}(\theta^*)}{\partial \theta}\middle| z_{jt}\right]'}_{h_E^*(z_{jt})}(\theta_0 - \theta^*)$$

back

## Comparison between approximated "optimal instruments" and interval instruments

- Both the approximated optimal instruments by [Reynaert and Verboven, 2014] and the interval instruments can be interpreted as approximations of the MPI.

# Comparison between approximated "optimal instruments" and interval instruments

- Both the approximated optimal instruments by [Reynaert and Verboven, 2014] and the interval instruments can be interpreted as approximations of the MPI.
    - [Reynaert and Verboven, 2014] takes fully advantage of the parametric assumption $f \in \mathcal{F}_0$ and should be more "precise" in parametric case

## Comparison between approximated "optimal instruments" and interval instruments

- Both the approximated optimal instruments by [Reynaert and Verboven, 2014] and the interval instruments can be interpreted as approximations of the MPI.
  - [Reynaert and Verboven, 2014] takes fully advantage of the parametric assumption $f \in \mathcal{F}_0$ and should be more "precise" in parametric case
  - requires good first stage estimates
  - what is estimated when the distribution is misspecified?
- Interval instruments:

# Comparison between approximated "optimal instruments" and interval instruments

- Both the approximated optimal instruments by [Reynaert and Verboven, 2014] and the interval instruments can be interpreted as approximations of the MPI.
    - [Reynaert and Verboven, 2014] takes fully advantage of the parametric assumption $f \in \mathcal{F}_0$ and should be more "precise" in parametric case
    - requires good first stage estimates
    - what is estimated when the distribution is misspecified?
- Interval instruments:
    - by construction, less sensitive to a poor first stage estimates
    - interval instruments can be derived without estimating the full model (with the simple logit specification)
    - possible interpretation of the estimates even when distribution is misspecified.

back

# Simulations estimation: set-up

- Setting close to [Dubé et al., 2012] and [Reynaert and Verboven, 2014] Implementation

  Indirect utility is given by

$$u_{ijt} = 2 + x_{ajt} + 1.5 x_{bjt} - 2 p_{jt} + x_{cjt} v_i + \xi_{jt} + \varepsilon_{ijt}, \quad \xi_{jt} \overset{iid}{\sim} \mathcal{N}(0,1), \quad \varepsilon_{ijt} \overset{iid}{\sim} EV(1)$$

  - $T \in \{50, 100, 200\}$ , $J = 12$
  - $x_a$, $x_b$, $x_c$ are normal and correlated
  - Price is endogenous $p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt}$

# Simulations estimation: set-up

- Setting close to [Dubé et al., 2012] and [Reynaert and Verboven, 2014] ▸ Implementation

  Indirect utility is given by

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt}, \quad \xi_{jt} \overset{iid}{\sim} \mathcal{N}(0,1), \quad \varepsilon_{ijt} \overset{iid}{\sim} EV(1)$$

  - $T \in \{50, 100, 200\}$ , $J = 12$
  - $x_a$, $x_b$, $x_c$ are normal and correlated
  - Price is endogenous $p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt}$

- Estimation is always done assuming the right specification: $f \in \mathcal{F}_0$

- $v_i \sim f$ and $f$ varies between a normal, mixture of normals, etc...

# Simulations estimation: set-up

- Setting close to [Dubé et al., 2012] and [Reynaert and Verboven, 2014] Implementation

  Indirect utility is given by

$$u_{ijt} = 2 + x_{ajt} + 1.5x_{bjt} - 2p_{jt} + x_{cjt}v_i + \xi_{jt} + \varepsilon_{ijt}, \quad \xi_{jt} \overset{iid}{\sim} \mathcal{N}(0,1), \quad \varepsilon_{ijt} \overset{iid}{\sim} EV(1)$$

  - $T \in \{50, 100, 200\}$ , $J = 12$
  - $x_a$, $x_b$, $x_c$ are normal and correlated
  - Price is endogenous $p_{jt} = 1 + \xi_{jt} + u_{jt} + \sum_{k=a}^{c} x_{kjt} + c_{1jt} + c_{2jt}$

- Estimation is always done assuming the right specification: $f \in \mathcal{F}_0$

- $v_i \sim f$ and $f$ varies between a normal, mixture of normals, etc...

- We consider different sets instruments:

  - Differentiation instruments [Gandhi and Houde, 2019]; "Optimal instrument " [Reynaert and Verboven, 2014]; Interval instruments

back

# Simulations: estimation of a mixture

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a mixture of 2 normal components:

$$f(v) = 0.25 f_L(v) + 0.75 f_H(v), \qquad f_L \sim \mathcal{N}(-2, 0.5) \quad f_H \sim \mathcal{N}(4, 0.5)$$

- We work under the null and try to estimate the parameters $(p_L, \beta_{3L}, \beta_{3H}, \sigma_{3L}, \sigma_{3H})$ using the different instruments

# Simulations: estimation of a mixture

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a mixture of 2 normal components:

$$f(v) = 0.25 f_L(v) + 0.75 f_H(v), \qquad f_L \sim \mathcal{N}(-2, 0.5) \quad f_H \sim \mathcal{N}(4, 0.5)$$

- We work under the null and try to estimate the parameters $(p_L, \beta_{3L}, \beta_{3H}, \sigma_{3L}, \sigma_{3H})$ using the different instruments

back

| | Instruments | Differentiation | | | | | "Optimal" | | | | | Interval Global | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ |
| Sample size | true | -2 | 0.5 | 4 | 0.5 | 0.25 | -2 | 0.5 | 4 | 0.5 | 0.25 | -2 | 0.5 | 4 | 0.5 | 0.25 |
| T=50, J=12 | bias | 0.204 | 0.175 | -0.024 | -0.043 | 0.025 | 0.074 | 0.057 | 0.026 | -0.11 | 0.01 | 0.015 | -0.005 | -0.045 | 0.006 | 0.004 |
| | $\sqrt{MSE}$ | 0.618 | 0.723 | 0.28 | 0.35 | 0.072 | 0.359 | 0.481 | 0.212 | 0.281 | 0.035 | 0.274 | 0.387 | 0.225 | 0.256 | 0.023 |
| T=100, J=12 | bias | 0.222 | 0.213 | 0.017 | -0.063 | 0.025 | 0.053 | 0.035 | 0.018 | -0.065 | 0.007 | 0 | -0.016 | -0.027 | 0.006 | 0.001 |
| | $\sqrt{MSE}$ | 0.569 | 0.689 | 0.248 | 0.304 | 0.067 | 0.278 | 0.398 | 0.154 | 0.21 | 0.028 | 0.132 | 0.268 | 0.156 | 0.2 | 0.005 |
| T=200, J=12 | bias | 0.166 | 0.147 | 0.008 | -0.049 | 0.017 | 0.072 | 0.104 | 0.033 | -0.074 | 0.01 | -0.006 | -0.027 | -0.015 | -0.001 | 0.001 |
| | $\sqrt{MSE}$ | 0.427 | 0.571 | 0.171 | 0.259 | 0.048 | 0.148 | 0.23 | 0.118 | 0.179 | 0.014 | 0.088 | 0.219 | 0.108 | 0.164 | 0.003 |

# Simulations: estimation of a mixture

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a mixture of 2 normal components:

$$f(v) = 0.25 f_L(v) + 0.75 f_H(v), \qquad f_L \sim \mathcal{N}(-2, 0.5) \quad f_H \sim \mathcal{N}(4, 0.5)$$

- We work under the null and try to estimate the parameters $(p_L, \beta_{3L}, \beta_{3H}, \sigma_{3L}, \sigma_{3H})$ using the different instruments

back

| | Instruments | Differentiation | | | | | "Optimal" | | | | | Interval Global | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ |
| Sample size | true | -2 | 0.5 | 4 | 0.5 | 0.25 | -2 | 0.5 | 4 | 0.5 | 0.25 | -2 | 0.5 | 4 | 0.5 | 0.25 |
| T=50, J=12 | bias | 0.204 | 0.175 | -0.024 | -0.043 | 0.025 | 0.074 | 0.057 | 0.026 | -0.11 | 0.01 | 0.015 | -0.005 | -0.045 | 0.006 | 0.004 |
| | $\sqrt{MSE}$ | 0.618 | 0.723 | 0.28 | 0.35 | 0.072 | 0.359 | 0.481 | 0.212 | 0.281 | 0.035 | 0.274 | 0.387 | 0.225 | 0.256 | 0.023 |
| T=100, J=12 | bias | 0.222 | 0.213 | 0.017 | -0.063 | 0.025 | 0.053 | 0.035 | 0.018 | -0.065 | 0.007 | 0 | -0.016 | -0.027 | 0.006 | 0.001 |
| | $\sqrt{MSE}$ | 0.569 | 0.689 | 0.248 | 0.304 | 0.067 | 0.278 | 0.398 | 0.154 | 0.21 | 0.028 | 0.132 | 0.268 | 0.156 | 0.2 | 0.005 |
| T=200, J=12 | bias | 0.166 | 0.147 | 0.008 | -0.049 | 0.017 | 0.072 | 0.104 | 0.033 | -0.074 | 0.01 | -0.006 | -0.027 | -0.015 | -0.001 | 0.001 |
| | $\sqrt{MSE}$ | 0.427 | 0.571 | 0.171 | 0.259 | 0.048 | 0.148 | 0.23 | 0.118 | 0.179 | 0.014 | 0.088 | 0.219 | 0.108 | 0.164 | 0.003 |

# Simulations: estimation of a mixture

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a mixture of 2 normal components:

$$f(v) = 0.25 f_L(v) + 0.75 f_H(v), \qquad f_L \sim \mathcal{N}(-2, 0.5) \quad f_H \sim \mathcal{N}(4, 0.5)$$

- We work under the null and try to estimate the parameters $(p_L, \beta_{3L}, \beta_{3H}, \sigma_{3L}, \sigma_{3H})$ using the different instruments

# Simulations: estimation of a mixture

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a mixture of 2 normal components:

$$f(v) = 0.25 f_L(v) + 0.75 f_H(v), \qquad f_L \sim \mathcal{N}(-2, 0.5) \quad f_H \sim \mathcal{N}(4, 0.5)$$

- We work under the null and try to estimate the parameters $(p_L, \beta_{3L}, \beta_{3H}, \sigma_{3L}, \sigma_{3H})$ using the different instruments

back

| | Instruments | Differentiation | | | | | "Optimal" | | | | | Interval Local | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ | $\beta_{3L}$ | $\sigma_{3L}$ | $\beta_{3H}$ | $\sigma_{3H}$ | $p_L$ |
| Sample size | true | -2 | 0.5 | 4 | 0.5 | 0.25 | -2 | 0.5 | 4 | 0.5 | 0.25 | -2 | 0.5 | 4 | 0.5 | 0.25 |
| T=50, J=12 | bias | 0.204 | 0.175 | -0.024 | -0.043 | 0.025 | 0.107 | 0.135 | 0.057 | -0.132 | 0.01 | -0.007 | -0.011 | -0.042 | 0.004 | 0.003 |
| | $\sqrt{MSE}$ | 0.618 | 0.723 | 0.28 | 0.35 | 0.072 | 0.342 | 0.49 | 0.223 | 0.307 | 0.028 | 0.241 | 0.334 | 0.212 | 0.242 | 0.017 |
| T=50, J=12 | bias | 0.222 | 0.213 | 0.017 | -0.063 | 0.025 | 0.016 | -0.002 | 0.009 | -0.05 | 0.003 | -0.001 | -0.001 | -0.028 | 0.009 | 0.001 |
| | $\sqrt{MSE}$ | 0.569 | 0.689 | 0.248 | 0.304 | 0.067 | 0.176 | 0.321 | 0.146 | 0.181 | 0.011 | 0.123 | 0.221 | 0.142 | 0.161 | 0.005 |
| T=50, J=12 | bias | 0.166 | 0.147 | 0.008 | -0.049 | 0.017 | 0.021 | 0.041 | 0.018 | -0.057 | 0.003 | 0.002 | -0.007 | -0.015 | 0.007 | 0.001 |
| | $\sqrt{MSE}$ | 0.427 | 0.571 | 0.171 | 0.259 | 0.048 | 0.148 | 0.23 | 0.118 | 0.179 | 0.014 | 0.091 | 0.173 | 0.098 | 0.121 | 0.003 |

# Simulations: estimation of a Gaussian

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a Gaussian: $v \sim \mathcal{N}(1.5, 0.5)$
- We work under the null and try to estimate the parameters $(\beta_0, \alpha, \beta_1, \beta_2, \beta_3, \sigma_3)$ using the different instruments

back

| | Instruments | Differentiation | | | | | | "Optimal" | | | | | | Interval global | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter | $\beta_0$ | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_3$ | $\beta_0$ | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_3$ | $\beta_0$ | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_3$ |
| Sample size | True | 2 | -2 | 1.5 | 1 | 1.5 | 0.5 | 2 | -2 | 1.5 | 1 | 1.5 | 0.5 | 2 | -2 | 1.5 | 1 | 1.5 | 0.5 |
| T=50, J=12 | bias | -0.16 | 0.032 | -0.03 | -0.028 | -0.032 | -0.003 | -0.09 | 0.018 | -0.016 | -0.014 | -0.018 | -0.003 | -0.15 | 0.03 | -0.028 | -0.026 | -0.03 | -0.004 |
| | $\sqrt{MSE}$ | 0.293 | 0.057 | 0.212 | 0.209 | 0.138 | 0.067 | 0.27 | 0.053 | 0.214 | 0.211 | 0.138 | 0.067 | 0.288 | 0.056 | 0.212 | 0.209 | 0.138 | 0.066 |
| T=50, J=12 | bias | -0.088 | 0.017 | -0.001 | 0 | -0.027 | 0.001 | -0.052 | 0.01 | 0.007 | 0.007 | -0.02 | 0.001 | -0.081 | 0.016 | 0.001 | 0.002 | -0.026 | 0.001 |
| | $\sqrt{MSE}$ | 0.199 | 0.039 | 0.146 | 0.146 | 0.101 | 0.045 | 0.189 | 0.037 | 0.148 | 0.147 | 0.099 | 0.047 | 0.197 | 0.039 | 0.146 | 0.145 | 0.1 | 0.044 |
| T=50, J=12 | bias | -0.038 | 0.007 | -0.012 | -0.012 | -0.004 | 0.002 | -0.017 | 0.003 | -0.006 | -0.007 | -0.001 | 0 | -0.032 | 0.006 | -0.009 | -0.01 | -0.004 | 0 |
| | $\sqrt{MSE}$ | 0.132 | 0.026 | 0.11 | 0.11 | 0.073 | 0.032 | 0.127 | 0.025 | 0.109 | 0.109 | 0.069 | 0.032 | 0.129 | 0.026 | 0.109 | 0.109 | 0.069 | 0.032 |

Figure 13: Estimation of a gaussian random coefficient

# Simulations: estimation of a Gaussian

- Consider a setting where $f \in \mathcal{F}_0$ and $f$ is a Gaussian: $v \sim \mathcal{N}(1.5, 0.5)$
- We work under the null and try to estimate the parameters $(\beta_0, \alpha, \beta_1, \beta_2, \beta_3, \sigma_3)$ using the different instruments

back

| | Instruments | Differentiation | | | | | | "Optimal" | | | | | | Interval local | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter | $\beta_0$ | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_3$ | $\beta_0$ | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_3$ | $\beta_0$ | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_3$ |
| Sample size | True | 2 | -2 | 1.5 | 1 | 1.5 | 0.5 | 2 | -2 | 1.5 | 1 | 1.5 | 0.5 | 2 | -2 | 1.5 | 1 | 1.5 | 0.5 |
| T=50, J=12 | bias | -0.16 | 0.032 | -0.03 | -0.028 | -0.032 | -0.003 | -0.09 | 0.018 | -0.016 | -0.014 | -0.018 | -0.003 | -0.15 | 0.03 | -0.028 | -0.026 | -0.03 | -0.001 |
| | $\sqrt{MSE}$ | 0.293 | 0.057 | 0.212 | 0.209 | 0.138 | 0.067 | 0.27 | 0.053 | 0.214 | 0.211 | 0.138 | 0.067 | 0.286 | 0.056 | 0.212 | 0.209 | 0.138 | 0.064 |
| T=50, J=12 | bias | -0.088 | 0.017 | -0.001 | 0 | -0.027 | 0.001 | -0.052 | 0.01 | 0.007 | 0.007 | -0.02 | 0.001 | -0.074 | 0.014 | -0.016 | -0.016 | -0.013 | 0.001 |
| | $\sqrt{MSE}$ | 0.199 | 0.039 | 0.146 | 0.146 | 0.101 | 0.045 | 0.189 | 0.037 | 0.148 | 0.147 | 0.099 | 0.047 | 0.185 | 0.036 | 0.151 | 0.152 | 0.099 | 0.044 |
| T=50, J=12 | bias | -0.038 | 0.007 | -0.012 | -0.012 | -0.004 | 0.002 | -0.017 | 0.003 | -0.006 | -0.007 | -0.001 | 0 | -0.032 | 0.006 | -0.009 | -0.01 | -0.004 | 0.001 |
| | $\sqrt{MSE}$ | 0.132 | 0.026 | 0.11 | 0.11 | 0.073 | 0.032 | 0.127 | 0.025 | 0.109 | 0.109 | 0.069 | 0.032 | 0.129 | 0.026 | 0.109 | 0.109 | 0.069 | 0.031 |

Figure 14: Estimation of a gaussian random coefficient

# Simulations implementation details

- For each setting, we estimate the model for 1000 replications
- For each replication, we choose 3 different starting values and we select the set of parameters with the lowest objective function
- Market shares are integrated using product rules
- Minimization is performed with nloptr ( algorithm: NLOPT-LD-LBFGS)
- Threshold for the outer loop: 1e-9. Threshold for the inner loop:1e-13
- We use squarem and a C++ implementation to speed up the contraction (we also parallelize over markets using 14 independent cores)

Back

# Simulations instruments details

- J test(1): differentiation instruments + exogenous characteristics (polynomial terms) + cost shifters (15 instruments/ degrees of overidentification:8)
- I test(1): First stage instruments from J test(1); Testing instruments are 7 interval instruments, points chosen as follows: $\{\hat{\mu}, (\hat{\mu} + k(max(0.25, \hat{\sigma})), k(max(0.25, \hat{\sigma}))\}$ ($for\ \ k = 1, 2, 3$)
- J test(2): First stage instruments are from J test(1); Second stage instruments are optimal instruments (approximation of $\mathbb{E}\left[\frac{\partial \rho_j^{-1}(s_t, x_{2t}, \lambda)}{\partial \lambda}\middle| z_t\right]$)+ exogenous characteristics (polynomial terms) + cost shifters (12 instruments)
- I test(2): First stage instruments from J test(2); Testing instruments are 7 interval instruments, points chosen as follows: $\{\hat{\mu}, (\hat{\mu} + k(max(0.25, \hat{\sigma})), k(max(0.25, \hat{\sigma}))\}$ ($for\ \ k = 1, 2, 3$) Back

# Outline

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity
  - The literature on this topic imposes restrictive assumptions on the dist of RC

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity
  - The literature on this topic imposes restrictive assumptions on the dist of RC
- How do the tools we develop in this paper help us?

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity
  - The literature on this topic imposes restrictive assumptions on the dist of RC
- How do the tools we develop in this paper help us?
  - **Model selection**: which variables to augment with a RC? size, horsepower, engine type, height, weight, fuel costs, CO2 emission, price

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity
  - The literature on this topic imposes restrictive assumptions on the dist of RC
- How do the tools we develop in this paper help us?
  - **Model selection**: which variables to augment with a RC? size, horsepower, engine type, height, weight, fuel costs, CO2 emission, price
  - **Specification test**: is the normality assumption on the RC on price rejected by the data?

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity
  - The literature on this topic imposes restrictive assumptions on the dist of RC
- How do the tools we develop in this paper help us?
  - **Model selection**: which variables to augment with a RC? size, horsepower, engine type, height, weight, fuel costs, CO2 emission, price
  - **Specification test**: is the normality assumption on the RC on price rejected by the data?
  - **Estimation:** need of "informative" instruments to estimate a more flexible distribution

# Introduction: Empirical application

- We want to study the effects on welfare and $C0_2$ emissions of different taxation schemes in the German car market
  - compare the performance of fuel tax and product tax
- The counterfactuals hinge on substitution patterns and the shape of preference heterogeneity
  - The literature on this topic imposes restrictive assumptions on the dist of RC
- How do the tools we develop in this paper help us?
  - **Model selection**: which variables to augment with a RC? size, horsepower, engine type, height, weight, fuel costs, CO2 emission, price
  - **Specification test**: is the normality assumption on the RC on price rejected by the data?
  - **Estimation:** need of "informative" instruments to estimate a more flexible distribution

# Data

- Most of the data was provided to us by Kevin Remmy (Mannheim)

- Data on state level new car registrations, publicly available by German Federal Motor Transport Authority (KBA) from 2012 to 2018.

  $\rightarrow$This gives us 112 markets defined by state-year pairs

- Data on car characteristics (General German Automobile Club): price, horsepower, engine type, size, weight, fuel costs, CO2 emission, ...

- We scraped cost shifters: distance to the plant, price of steel, average cost of labor in assembly country, exchange rates between Germany and production country

- We aggregate by Brand, Model, FuelType, Body and remove very low shares $\rightarrow$ 33,760 observations

# Summary statistics

Figure 15: Summary Statistics (Sales weighted)

| | Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| **Diesel** | | | | | | | |
| Price/income | 0.74 | 0.72 | 0.73 | 0.72 | 0.71 | 0.69 | 0.68 |
| Size (m2) | 8.31 | 8.31 | 8.32 | 8.36 | 8.42 | 8.48 | 8.53 |
| Horsepower (kW/100) | 1.09 | 1.07 | 1.11 | 1.11 | 1.14 | 1.16 | 1.21 |
| Fuel cost (euros/100km) | 7.90 | 7.18 | 6.63 | 5.53 | 4.94 | 5.25 | 5.83 |
| Fuel cons. (Lt./100km) | 5.19 | 4.98 | 4.89 | 4.73 | 4.61 | 4.61 | 4.71 |
| CO2 emission (g/km) | 136.19 | 130.50 | 127.69 | 123.58 | 120.42 | 120.49 | 123.27 |
| Nb. of products/market | 133 | 138 | 146 | 150 | 151 | 149 | 143 |
| **Gasoline** | | | | | | | |
| Price/income | 0.46 | 0.46 | 0.46 | 0.46 | 0.46 | 0.45 | 0.43 |
| Size (m2) | 7.23 | 7.27 | 7.28 | 7.30 | 7.36 | 7.46 | 7.53 |
| Horsepower (kW/100) | 0.79 | 0.78 | 0.80 | 0.82 | 0.85 | 0.88 | 0.91 |
| Fuel cost (euros/100km) | 9.48 | 8.61 | 8.11 | 7.27 | 6.69 | 7.06 | 7.40 |
| Fuel cons. (Lt./100km) | 5.76 | 5.47 | 5.40 | 5.31 | 5.25 | 5.34 | 5.38 |
| CO2 emission (g/km) | 135.80 | 128.18 | 125.27 | 122.89 | 121.22 | 122.86 | 123.26 |
| Nb. of products/market | 157 | 171 | 179 | 185 | 186 | 193 | 188 |

Note: Provided statistics are sales weighted averages across products. Total number of markets (State*Year) is 112

# Results logit and nested logit

| | OLS | IV | |
|---|---|---|---|
| Price/income | −0.542*** | −2.448*** | −2.410*** |
| | (0.035) | (0.118) | (0.048) |
| log(within market shares) | | | 0.432*** |
| | | | (0.006) |
| Fuel Cost | −0.127*** | −0.136*** | −0.089*** |
| | (0.006) | (0.006) | (0.004) |
| Size($m^2$) | −0.207*** | −0.101*** | 0.038*** |
| | (0.010) | (0.012) | (0.008) |
| Horsepower(KW/100) | 0.331*** | 1.431*** | 0.985*** |
| | (0.030) | (0.072) | (0.028) |
| Foreign | −0.568*** | −0.577*** | −0.466*** |
| | (0.018) | (0.019) | (0.012) |
| Height(m) | 0.335*** | 0.759*** | 0.323*** |
| | (0.075) | (0.082) | (0.048) |
| Gasoline | 0.620*** | 0.499*** | 0.260*** |
| | (0.019) | (0.021) | (0.013) |
| Constant | −8.678*** | −10.146*** | −7.054*** |
| | (0.150) | (0.178) | (0.099) |
| Market FE | *Yes* | *Yes* | *Yes* |
| Observations | 39,888 | 39,888 | 39,888 |
| $R^2$ | 0.372 | 0.326 | 0.746 |

# Variable selection

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 539.0 | 16.9 |
| I test all | 3292.0 | 47.4 |
| I test price | 826.2 | 42.1 |
| I test fuel cost | 766.8 | 42.1 |
| I test size | 1334.4 | 42.1 |
| I test horsepower | 781.9 | 42.1 |
| I test gazoline | 28.5 | 42.1 |
| I test Foreign | 177.2 | 42.1 |
| I test Height | 411.1 | 42.1 |

# Variable selection

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 539.0 | 16.9 |
| I test all | 3292.0 | 47.4 |
| I test price | 826.2 | 42.1 |
| I test fuel cost | 766.8 | 42.1 |
| I test size | 1334.4 | 42.1 |
| I test horsepower | 781.9 | 42.1 |
| I test gazoline | 28.5 | 42.1 |
| I test Foreign | 177.2 | 42.1 |
| I test Height | 411.1 | 42.1 |

# Variable selection

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 539.0 | 16.9 |
| I test all | 3292.0 | 47.4 |
| I test price | 826.2 | 42.1 |
| I test fuel cost | 766.8 | 42.1 |
| I test size | 1334.4 | 42.1 |
| I test horsepower | 781.9 | 42.1 |
| I test gazoline | 28.5 | 42.1 |
| I test Foreign | 177.2 | 42.1 |
| I test Height | 411.1 | 42.1 |

# Variable selection

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 539.0 | 16.9 |
| I test all | 3292.0 | 47.4 |
| I test price | 826.2 | 42.1 |
| I test fuel cost | 766.8 | 42.1 |
| I test size | 1334.4 | 42.1 |
| I test horsepower | 781.9 | 42.1 |
| I test gazoline | 28.5 | 42.1 |
| I test Foreign | 177.2 | 42.1 |
| I test Height | 411.1 | 42.1 |

# Variable selection

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 539.0 | 16.9 |
| I test all | 3292.0 | 47.4 |
| I test price | 826.2 | 42.1 |
| I test fuel cost | 766.8 | 42.1 |
| I test size | 1334.4 | 42.1 |
| I test horsepower | 781.9 | 42.1 |
| I test gazoline | 28.5 | 42.1 |
| I test Foreign | 177.2 | 42.1 |
| I test Height | 411.1 | 42.1 |

# Results standard BLP

Figure 16: Standard BLP estimation

| | estimate | standard error |
|---|---|---|
| Price/income | -1.50e+00 | 1.16e-01 |
| sd Price | 3.30e-01 | 4.36e-02 |
| Fuel Cost | -1.22e-01 | 5.69e-03 |
| sd Fuel Cost | 1.11e-07 | 1.19e-22 |
| Size(m^2) | -1.83e+00 | 6.51e-02 |
| sd size | 9.40e-01 | 6.50e-02 |
| Horsepower(KW/100) | 4.19e-01 | 6.90e-02 |
| sd Horsepower | 4.69e-01 | 2.97e-02 |
| Foreign | -6.18e-01 | 2.02e-02 |
| Height(m) | 1.97e-01 | 7.59e-02 |
| Gasoline | 5.19e-01 | 2.04e-02 |
| constant | -3.88e+00 | 1.77e-01 |

BLP with random coefficients on price, fuel cost, power, size

# Results standard BLP

Figure 16: Standard BLP estimation

| | estimate | standard error |
|---|---|---|
| Price/income | -1.50e+00 | 1.16e-01 |
| sd Price | 3.30e-01 | 4.36e-02 |
| Fuel Cost | -1.22e-01 | 5.69e-03 |
| sd Fuel Cost | 1.11e-07 | 1.19e-22 |
| Size(m^2) | -1.83e+00 | 6.51e-02 |
| sd size | 9.40e-01 | 6.50e-02 |
| Horsepower(KW/100) | 4.19e-01 | 6.90e-02 |
| sd Horsepower | 4.69e-01 | 2.97e-02 |
| Foreign | -6.18e-01 | 2.02e-02 |
| Height(m) | 1.97e-01 | 7.59e-02 |
| Gasoline | 5.19e-01 | 2.04e-02 |
| constant | -3.88e+00 | 1.77e-01 |

BLP with random coefficients on price, fuel cost, power, size

# Results standard BLP

Figure 16: Standard BLP estimation

| | estimate | standard error |
|---|---|---|
| Price/income | -1.50e+00 | 1.16e-01 |
| sd Price | 3.30e-01 | 4.36e-02 |
| Fuel Cost | -1.22e-01 | 5.69e-03 |
| sd Fuel Cost | 1.11e-07 | 1.19e-22 |
| Size(m^2) | -1.83e+00 | 6.51e-02 |
| sd size | 9.40e-01 | 6.50e-02 |
| Horsepower(KW/100) | 4.19e-01 | 6.90e-02 |
| sd Horsepower | 4.69e-01 | 2.97e-02 |
| Foreign | -6.18e-01 | 2.02e-02 |
| Height(m) | 1.97e-01 | 7.59e-02 |
| Gasoline | 5.19e-01 | 2.04e-02 |
| constant | -3.88e+00 | 1.77e-01 |

BLP with random coefficients on price, fuel cost, power, size

# Specification test

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 2390.3 | 16.9 |
| I test all | 1388.7 | 37.7 |
| I test price | 112.7 | 42.1 |
| I test fuel cost | 86.3 | 42.1 |
| I test size | 246.2 | 42.1 |
| I test horsepower | 101.6 | 42.1 |
| I test gazoline | 89.0 | 42.1 |
| I test Foreign | 95.8 | 42.1 |
| I test Height | 87.4 | 42.1 |

# Specification test

| covariate/test | Value statistic | critical value |
|---|---|---|
| J test | 2390.3 | 16.9 |
| I test all | 1388.7 | 37.7 |
| I test price | 112.7 | 42.1 |
| I test fuel cost | 86.3 | 42.1 |
| I test size | 246.2 | 42.1 |
| I test horsepower | 101.6 | 42.1 |
| I test gazoline | 89.0 | 42.1 |
| I test Foreign | 95.8 | 42.1 |
| I test Height | 87.4 | 42.1 |

# Results logit and nested logit

Figure 17: Estimation results - Logit and Nested Logit

| | *OLS* | | *IV* | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Price/income | −0.442*** | −2.338*** | −3.103*** | −2.372*** | −2.992*** |
| | (0.043) | (0.126) | (0.155) | (0.065) | (0.068) |
| log(within market shares) | | | | 0.410*** | 0.466*** |
| | | | | (0.007) | (0.007) |
| Fuel Cost | −0.171*** | 0.002 | −0.164*** | 0.007** | −0.078*** |
| | (0.007) | (0.005) | (0.007) | (0.003) | (0.005) |
| Size($m^2$) | −0.239*** | −0.189*** | −0.089*** | −0.023** | 0.067*** |
| | (0.011) | (0.013) | (0.014) | (0.009) | (0.009) |
| Horsepower(KW/100) | 0.312*** | 1.028*** | 1.817*** | 0.750*** | 1.249*** |
| | (0.034) | (0.069) | (0.091) | (0.033) | (0.036) |
| Foreign | −0.465*** | −0.458*** | −0.415*** | −0.407*** | −0.376*** |
| | (0.019) | (0.020) | (0.020) | (0.013) | (0.012) |
| Height(m) | 0.701*** | 0.358*** | 1.017*** | 0.118** | 0.452*** |
| | (0.081) | (0.084) | (0.087) | (0.054) | (0.051) |
| Gasoline | 0.602*** | 0.003 | 0.402*** | −0.073*** | 0.114*** |
| | (0.022) | (0.023) | (0.025) | (0.014) | (0.016) |
| Constant | −8.255*** | −9.588*** | −10.011*** | −7.058*** | −7.080*** |
| | (0.173) | (0.188) | (0.207) | (0.121) | (0.110) |
| State FE & Year FE | ✓ | | ✓ | | ✓ |
| Observations | 33,760 | 33,760 | 33,760 | 33,760 | 33,760 |
| $R^2$ | 0.372 | 0.290 | 0.301 | 0.701 | 0.757 |

# Feasible approximation of the MPI: Riemann sum

- **Integral approximation**: we approximate directly the integral in which $f$ appears with a finite Riemann sum

$$\int_{\mathbb{R}^{K_2}} k(x_{2t}'v, s_t, \mathcal{F}_0, \theta_0)(\mathbf{f_0}(\mathbf{v}|\lambda_\mathbf{0}) - f(v))dv \approx \sum_{k=1}^{L} \underbrace{\frac{v_L - v_0}{L} h(x_{2t}'v_k, s_t, \mathcal{F}_0, \theta_0)}_{known} \underbrace{\alpha_k}_{unknown}$$

with

- $\alpha_k = f(v_k) - f_0(v_k|\lambda)$.

-L: number of points in the Riemann sum: $\{v_k\}_{k=1,\dots,L}$

-The approximation of the MPI is a linear combination of known terms

$\rightarrow$ Each element corresponds to one instrument $\rightarrow$ interval instruments

back

# Principle behind interval instruments

- Decompose the error term $\xi_{jt}(\mathcal{F}_0, \theta_0)$:

$$\xi_t(\mathcal{F}_0, \theta_0) = \xi_t(f, \theta) + \underbrace{(id - M)(\Delta(s_t, x_{2t}, \mathcal{F}_0, f))}_{\text{correction term due to misspecification}}$$

where $\Delta(s_t, x_{2t}, \mathcal{F}_0, f) = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0)) - \rho^{-1}(s_t, x_{2t}, f)$ and

$$M(\cdot) = x_{1t}' \left( \mathbb{E}\left[ \sum_j x_{1jt} h_E(z_{jt})' \right] W \mathbb{E}\left[ \sum_j h_E(z_{jt}) x_{1jt}' \right] \right)^{-1} \left( \mathbb{E}\left[ \sum_j x_{1jt} h_E(z_{jt})' \right] W \mathbb{E}\left[ \sum_j h_E(z_{jt}) \cdot \right] \right)$$

# Principle behind interval instruments

- Decompose the error term $\xi_{jt}(\mathcal{F}_0, \theta_0)$:

$$\xi_t(\mathcal{F}_0, \theta_0) = \xi_t(f, \theta) + \underbrace{(id - M)(\Delta(s_t, x_{2t}, \mathcal{F}_0, f))}_{\text{correction term due to misspecification}}$$

where $\Delta(s_t, x_{2t}, \mathcal{F}_0, f) = \rho^{-1}(s_t, x_{2t}, f_0(\cdot|\lambda_0)) - \rho^{-1}(s_t, x_{2t}, f)$ and

$$M(\cdot) = x'_{1t}\left(\mathbb{E}\left[\sum_j x_{1jt} h_E(z_{jt})'\right] W \mathbb{E}\left[\sum_j h_E(z_{jt}) x'_{1jt}\right]\right)^{-1}\left(\mathbb{E}\left[\sum_j x_{1jt} h_E(z_{jt})'\right] W \mathbb{E}\left[\sum_j h_E(z_{jt})\ \cdot\ \right]\right)$$

1 **Approximate the correction term by taking a first order "expansion" of $\rho_j^{-1}(s_t, x_{2t}, f) = \rho_j^{-1}(s_{0t}, x_{2t}, f_0(\cdot|\lambda_0))$ around $s_t$**

$$\Delta_j(s_t, x_{2t}, \mathcal{F}_0, f) = -e'_j\left(\frac{\partial \rho(\delta_t^0, x_{2t}, f_0(.|\lambda_0))}{\partial \delta}\right)^{-1} \int_{\mathbb{R}^{K_2}} \frac{exp(\delta_t^0 + x_{2t} v)}{1 + \sum_{k=1}^J \exp\left\{\delta_{kt}^0 + x'_{2jk} v\right\}} (f(v) - f_0(v|\lambda_0)) dv + \mathcal{R}_0 \quad \text{where}$$

$\mathcal{R}_0 = o\left(\int |f_0(.|\lambda_0) - f(v)| dv\right)$, $\delta_t^0 = \rho^{-1}(s_t, x_{2t}, f_0(.|\lambda_0))$

back

# Principle behind interval instruments

2 **Approximate the integral which appears in the correction term approximation with a Riemann sum**

$$\int_{\mathbb{R}} \frac{\exp\left(\delta_{jt}^0 + x_{2jt} v\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt}^0 + x_{2kt} v\right)} (f_0(v|\lambda_0) - f(v)) dv \approx \frac{v_1 - v_0}{L} \sum_{l=1}^{L} \underbrace{\frac{\exp\left\{\delta_{jt}^0 + x_{2jt} v_k\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_{kt} + x_{2kt} v_l\right\}}}_{known} \underbrace{\alpha_k}_{unknown}$$

where $\alpha_k = f_0(v_k|\lambda_0) - f(v_l)$ is unknown and $\{v_l\}_{l=1,\dots,L}$ points chosen on a grid over the support of $f_0(\cdot|\lambda_0) - f(\cdot)$

# Principle behind interval instruments

2 **Approximate the integral which appears in the correction term approximation with a Riemann sum**

$$\int_{\mathbb{R}} \frac{\exp\left(\delta^0_{jt} + x_{2jt}v\right)}{1+\sum_{k=1}^{J}\exp\left(\delta^0_{kt} + x_{2kt}v\right)}(f_0(v|\lambda_0) - f(v))dv \approx \frac{v_1 - v_0}{L}\sum_{l=1}^{L} \underbrace{\frac{\exp\left\{\delta^0_{jt} + x_{2jt}v_k\right\}}{1+\sum_{k=1}^{J}\exp\left\{\delta_{kt} + x_{2kt}v_l\right\}}}_{known} \underbrace{\alpha_k}_{unknown}$$

where $\alpha_k = f_0(v_k|\lambda_0) - f(v_l)$ is unknown and $\{v_l\}_{l=1,\dots,L}$ points chosen on a grid over the support of $f_0(\cdot|\lambda_0) - f(\cdot)$

3 **Exogenize the characteristics and** $\delta^0_t$, can be done in 2 ways

- Project $x$ on the instruments $h_E(z)$ and consider only the exogenous part of $\delta^0$, ie $\delta^0_{jt} = x'_{1jt}\beta_0$ as in [Reynaert and Verboven, 2014]
- Estimate the expectation of $(id - M)\Delta(\mathcal{F}_0, f)$ conditional on $z$ using a Sieve estimator, which in practice is not better than the 1st option

back

# Interval instruments implementation

1. Given $(\mathcal{F}_0, \hat{W}, h_E)$ obtain a BLP estimator with the method of your choice

## Interval instruments implementation

1. Given $(\mathcal{F}_0, \hat{W}, h_E)$ obtain a BLP estimator with the method of your choice
2. Exogenize $(x_1, x_2)$ by projecting them on $h_E(z)$

## Interval instruments implementation

1. Given $(\mathcal{F}_0, \hat{W}, h_E)$ obtain a BLP estimator with the method of your choice
2. Exogenize $(x_1, x_2)$ by projecting them on $h_E(z)$
3. Interval Instruments $\hat{h}^*_D(z)$ write

$$\hat{h}^*_D(z) = (I_{J\times T} - x_1(x_1' h_E(z)\hat{W} h_E(z)' x_1)^{-1}(x_1' h_E(z)\hat{W} h_E(z)')\hat{\Delta}_L$$

$$\hat{\Delta}_{jt,L} = \left\{ e_j' \left( \frac{\partial \rho(x_{1t}\hat{\beta}, x_{2t}, f_0(.|\hat{\lambda}))}{\partial \delta} \right)^{-1} \hat{\eta}_{t,l} \right\}_{l=1,\dots,L}$$

where $e_j = (0, 0, \dots, \underbrace{1}_{j_t hterm}, \dots, 0, 0)$, $\hat{\beta} = \hat{\beta}(\mathcal{F}_0, \hat{W}, h_E)$ and $\hat{\lambda} = \hat{\lambda}(\mathcal{F}_0, \hat{W}, h_E)$ are estimators of $\beta_0$ and $\lambda_0$, and

$$\hat{\eta}_{jt,l} = \frac{exp(x_{1jt}'\hat{\beta} + x_{2kt}' v_l)}{1 + \sum_{k=1}^{J} exp(x_{1kt}'\hat{\beta} + x_{2jt}' v_l)}$$

for some $(v_l)_{l=1}^L$ which are $L$ points taken in the support of $f_0(\cdot|\hat{\lambda})$. back

# Construction of the instruments in practice

- **Objective: Approximate** $h_D^*(z_{jt}) = \mathbb{E}(\Delta_{jt}(\mathcal{F}_0, f)|z_{jt})$

$\Rightarrow$ We build a vector of $L$ **interval instruments** $\hat{h}_D^*(z_{jt})$ using a first order approximation of $\xi_{jt}(\mathcal{F}_0, f) - \xi_{jt}$ and a guess on the support of $f$ ▸ Principle ▸ Implementation

# Construction of the instruments in practice

- **Objective: Approximate** $h_D^*(z_{jt}) = \mathbb{E}(\Delta_{jt}(\mathcal{F}_0, f)|z_{jt})$

$\Rightarrow$ We build a vector of $L$ **interval instruments** $\hat{h}_D^*(z_{jt})$ using a first order approximation of $\xi_{jt}(\mathcal{F}_0, f) - \xi_{jt}$ and a guess on the support of $f$ ▸ Principle ▸ Implementation

- **We prove that under certain conditions a linear combination of $\hat{h}_D^*(z_{jt})$ approximates $h_D^*$, and when $L$ is large that they have similar slopes** ▸ Sketch proof

- To prevent many / weak IV problems, $L$ cannot be too large in practice

- For similar reasons, $\hat{h}_D^*$ can be used for estimation with great effect ▸ Details

# Interval instruments sketch proof

$\rightarrow$ We show that there exists some $\alpha \in \mathbb{R}^L_*$ and some IV vector $\hat{h}^*_D(z_{jt})$ such that

$$\lim_{L\rightarrow\infty} \alpha' \hat{h}^*_D(z_{jt}) = h^*_D(z_{jt})$$

$\rightarrow$ In other words there exists a linear combination of $\hat{h}^*_D$ which approximates $h^*_D$

$\rightarrow$ A linear combination of $\hat{h}^*_D$ gives a smaller slope than using $\hat{h}^*_D$, ie $C_{\alpha' \hat{h}^*_D} \leqslant C_{\hat{h}^*_D}$, therefore

$$\lim_{L\rightarrow\infty} C_{\alpha' \hat{h}^*_D} = \lim_{L\rightarrow\infty} C_{\hat{h}^*_D} = C_{h^*_D}$$

Back

# References I


Andrews, Isaiah Gentzkow, M. and Shapiro, J. M. (2017).
Measuring the sensitivity of parameter estimates to estimation moments.
*Measuring the Sensitivity of Parameter Estimates to Estimation Moments*, 132:1153–1592.

Bahadur, R. R. (1960).
Stochastic comparison of tests.
*The Annals of Mathematical Statistics*, 31(2):276–295.

Berry, S., Levinsohn, J., and Pakes, A. (1995).
Automobile prices in market equilibrium.
*Econometrica: Journal of the Econometric Society*, pages 841–890.

Berry, S. T. (1994).
Estimating discrete-choice models of product differentiation.
*The RAND Journal of Economics*, pages 242–262.

Caputo, V., Scarpa, R., Nayga Jr, R. M., and Ortega, D. L. (2018).
Are preferences for food quality attributes really normally distributed? an analysis using flexible mixing distributions.
*Journal of choice modelling*, 28:10–27.

# References II


Compiani, G. (2018).

Nonparametric demand estimation in differentiated products markets.

*Available at SSRN 3134152.*

Conlon, C. and Gortmaker, J. (2019).

Best practices for differentiated products demand estimation with pyblp.

Dubé, J.-P., Fox, J. T., and Su, C.-L. (2012).

Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation.

*Econometrica*, 80(5):2231–2267.

Fosgerau, M. and Hess, S. (2009).

A comparison of methods for representing random taste heterogeneity in discrete choice models.

*European Transport-Trasporti Europei*, 42:1–25.

Fox, J. T. and Gandhi, A. (2016).

Nonparametric identification and estimation of random coefficients in multinomial choice models.

*The RAND Journal of Economics*, 47(1):118–139.

# References III


Fox, J. T., il Kim, K., Ryan, S. P., and Bajari, P. (2012).
The random coefficients logit model is identified.
*Journal of Econometrics*, 166(2):204–212.

Fox, J. T., Kim, K. I., Ryan, S. P., and Bajari, P. (2011).
A simple estimator for the distribution of random coefficients.
*Quantitative Economics*, 2(3):381–418.

Gandhi, A. and Houde, J.-F. (2019).
Measuring substitution patterns in differentiated products industries.
Technical report, National Bureau of Economic Research.

Gentzkow, M. and Shapiro, J. M. (2006).
Media bias and reputation.
*Journal of political Economy*, 114(2):280–316.

# References IV


Geweke, J. (1981).
The approximate slopes of econometric tests.
*Econometrica: Journal of the Econometric Society*, pages 1427–1442.

Hui, Müller, and Welsh (2021).
Random Effects Misspecification Can Have Severe Consequences for Random Effects Inference in Linear Mixed Models.
*International Statistical Review*, 86(1):186–206.

Knittel, C. R. and Metaxoglou, K. (2008).
Estimation of random coefficient demand models: Challenges, difficulties and warnings.
Technical report, National Bureau of Economic Research.

Lee, J. and Seo, K. (2015).
A computationally fast estimator for random coefficients logit demand models using aggregate data.
*The RAND Journal of Economics*, 46(1):86–102.

Lu, Z., Shi, X., and Tao, J. (2019).
Semi-nonparametric estimation of random coefficient logit model for aggregate demand.

# References V

McCulloch, C. E. and Neuhaus, J. M. (2011).
Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter.
*Statistical Science*, 26(3).

McFadden, D. and Train, K. (2000).
Mixed MNL models for discrete response.
*Journal of Applied Econometrics*, 15(5):447–470.

Nevo, A. (2000).
Mergers with differentiated products: The case of the ready-to-eat cereal industry.
*The RAND Journal of Economics*, pages 395–421.

Petrin, A. (2002).
Quantifying the benefits of new products: The case of the minivan.
*Journal of political Economy*, 110(4):705–729.

# References VI


Reynaert, M. and Verboven, F. (2014).
Improving the performance of random coefficients demand models: the role of optimal instruments.
*Journal of Econometrics*, 179(1):83–98.

Skrainka, B. S. and Judd, K. L. (2011).
High performance quadrature rules: How numerical integration affects a popular model of product differentiation.
*Available at SSRN 1870703*.

Train, K. (2016).
Mixed logit with a flexible mixing distribution.
*Journal of choice modelling*, 19:40–53.

Vij, A. and Krueger, R. (2017).
Random taste heterogeneity in discrete choice models: Flexible nonparametric finite mixture distributions.
*Transportation Research Part B: Methodological*, 106:76–101.

Wang, A. (2021).
Sieve blp: A semi-nonparametric model of demand for differentiated products.
*Available at SSRN 3569077*.

# References VII


White, H. (1982).
Maximum likelihood estimation of misspecified models.
*Econometrica: Journal of the econometric society*, pages 1–25.