Estimating the Price Elasticity of Gasoline Demand in Correlated Random Coefficient Models with Endogeneity

Michael Bates
UC-Riverside

Seolah Kim
Albion College

June, 2023

Motivation

- The price elasticity of gasoline demand is important for researchers modeling automotive and energy markets and for policy makers in urban planning and most importantly for mitigating climate change

Motivation

- The price elasticity of gasoline demand is important for researchers modeling automotive and energy markets and for policy makers in urban planning and most importantly for mitigating climate change
- Estimation is difficult due to:
    - Endogeneity: Simultaneous supply and demand forces

Motivation

- The price elasticity of gasoline demand is important for researchers modeling automotive and energy markets and for policy makers in urban planning and most importantly for mitigating climate change
- Estimation is difficult due to:
  - Endogeneity: Simultaneous supply and demand forces
  - Seasonal prices and quantities trend over time

Motivation

- The price elasticity of gasoline demand is important for researchers modeling automotive and energy markets and for policy makers in urban planning and most importantly for mitigating climate change
- Estimation is difficult due to:
  - Endogeneity: Simultaneous supply and demand forces
  - Seasonal prices and quantities trend over time
  - Heterogeneity: locations differ in taxes, regulation, infrastructure, macroeconomic climate (Wadud et al., 2010; Frondel et al., 2012; Blundell et al., 2012; Hausman and Newey, 2016)

Existing strategies for endogeneity

- We have an instrument! We follow the literature in instrumenting prices with state fuel taxes (Davis and Kilian, 2011; Blundell et al., 2012; Hausman and Newey, 2016; Coglianese et al., 2017; Hoderlein and Vanhems, 2018)

Existing strategies for endogeneity

- We have an instrument! We follow the literature in instrumenting prices with state fuel taxes (Davis and Kilian, 2011; Blundell et al., 2012; Hausman and Newey, 2016; Coglianese et al., 2017; Hoderlein and Vanhems, 2018)
- Given panel structure, use P2SLS or FEIV with two-way fixed effects to get a LATE?

Existing strategies for endogeneity

- We have an instrument! We follow the literature in instrumenting prices with state fuel taxes (Davis and Kilian, 2011; Blundell et al., 2012; Hausman and Newey, 2016; Coglianese et al., 2017; Hoderlein and Vanhems, 2018)
- Given panel structure, use P2SLS or FEIV with two-way fixed effects to get a LATE?
- Even with exogenous regressors, recent work explores issues from heterogeneous effects in two-way fixed effects models

[De Chaisemartin and d'Haultfoeuille (2020); Sun and Abraham (2021); Wooldridge (2021); Goldsmith-Pinkham et al. (2022); #metricstotheface]

Complications from heterogeneity

Simple model (Murtazashvili and Wooldridge, 2008):

- $y_{ij} = \alpha_i + x_{ij}(\beta + d_i) + e_{ij}$
- $i = 1, ..., N$ indexes cluster, $j = 1, .., T$ indexes observations within cluster, have instrument z_{ij}

Complications from heterogeneity

Simple model (Murtazashvili and Wooldridge, 2008):

- $y_{ij} = \alpha_i + x_{ij}(\beta + d_i) + e_{ij}$
- $i = 1, ..., N$ indexes cluster, $j = 1, .., T$ indexes observations within cluster, have instrument z_{ij}

$$\hat{\beta}_{FEIV} = \beta + \left(\sum_{i=1}^{N} \sum_{j=1}^{T} \ddot{z}'_{ij} \ddot{x}_{ij} \right)^{-1} \left[\sum_{i=1}^{N} \sum_{j=1}^{T} \ddot{z}'_{ij} \ddot{x}_{ij} d_i + \sum_{i=1}^{N} \sum_{j=1}^{T} \ddot{z}'_{ij} \ddot{e}_{ij} \right]$$

- Consistency of FEIV requires:

$$E[(\ddot{z}'_{ij} \ddot{x}_{ij})^{-1} \ddot{z}'_{ij} \ddot{x}_{ij} d_i] \neq 0$$

Complications from heterogeneity

Simple model (Murtazashvili and Wooldridge, 2008):

- $y_{ij} = \alpha_i + x_{ij}(\beta + d_i) + e_{ij}$
- $i = 1, ..., N$ indexes cluster, $j = 1, .., T$ indexes observations within cluster, have instrument z_{ij}

$$\hat{\beta}_{FEIV} = \beta + \left(\sum_{i=1}^{N} \sum_{j=1}^{T} \ddot{z}_{ij}' \ddot{x}_{ij} \right)^{-1} \left[\sum_{i=1}^{N} \sum_{j=1}^{T} \ddot{z}_{ij}' \ddot{x}_{ij} d_i + \sum_{i=1}^{N} \sum_{j=1}^{T} \ddot{z}_{ij}' \ddot{e}_{ij} \right]$$

- Consistency of FEIV requires:

$$E[(\ddot{z}_{ij}' \ddot{x}_{ij})^{-1} \ddot{z}_{ij}' \ddot{x}_{ij} d_i] \neq 0$$

Violations:

- A LATE different from the ATE exists
- Even in a population of compliers, correlations between the strength of IV and heterogeneous responsiveness leads to inconsistency

Correlated random coefficients–existing approaches

- Unobserved effect heterogeneity seems realistic in many contexts but is often not modeled

Correlated random coefficients–existing approaches

- Unobserved effect heterogeneity seems realistic in many contexts but is often not modeled
- Some existing estimators assume homogeneous or uncorrelated first-stage across individuals (Murtazashvili and Wooldridge, 2008, 2016; Laage, 2019)

Correlated random coefficients–existing approaches

- Unobserved effect heterogeneity seems realistic in many contexts but is often not modeled
- Some existing estimators assume homogeneous or uncorrelated first-stage across individuals (Murtazashvili and Wooldridge, 2008, 2016; Laage, 2019)
- Fernández-Val and Lee (2013) adopt a natural approach of estimating cluster-specific coefficients using GMM per cluster and averages over them

Correlated random coefficients–existing approaches

- Unobserved effect heterogeneity seems realistic in many contexts but is often not modeled
- Some existing estimators assume homogeneous or uncorrelated first-stage across individuals (Murtazashvili and Wooldridge, 2008, 2016; Laage, 2019)
- Fernández-Val and Lee (2013) adopt a natural approach of estimating cluster-specific coefficients using GMM per cluster and averages over them
- The Fernández-Val and Lee (2013) estimator does not, however, handle two-way fixed effects models (and may have difficulty with other exogenous regressors)

Per-Cluster Instrumental Variable Approach: New

- We show how to use cross-sectional variation to handle two-way fixed effects and exogenous regressors and within variation to obtain cluster-specific effects

Per-Cluster Instrumental Variable Approach: New

- We show how to use cross-sectional variation to handle two-way fixed effects and exogenous regressors and within variation to obtain cluster-specific effects
- Per-Cluster Instrumental Variables (PCIV) averages over these to consistently estimate Population Average Effects or LATEs

Per-Cluster Instrumental Variable Approach: New

- We show how to use cross-sectional variation to handle two-way fixed effects and exogenous regressors and within variation to obtain cluster-specific effects
- Per-Cluster Instrumental Variables (PCIV) averages over these to consistently estimate Population Average Effects or LATEs
- Provide standard errors for robust inference

Per-Cluster Instrumental Variable Approach: New

- We show how to use cross-sectional variation to handle two-way fixed effects and exogenous regressors and within variation to obtain cluster-specific effects
- Per-Cluster Instrumental Variables (PCIV) averages over these to consistently estimate Population Average Effects or LATEs
- Provide standard errors for robust inference
- Monte-Carlo finite sample simulations

Per-Cluster Instrumental Variable Approach: New

- We show how to use cross-sectional variation to handle two-way fixed effects and exogenous regressors and within variation to obtain cluster-specific effects
- Per-Cluster Instrumental Variables (PCIV) averages over these to consistently estimate Population Average Effects or LATEs
- Provide standard errors for robust inference
- Monte-Carlo finite sample simulations
- Apply the estimators to estimate the price elasticity of demand for gasoline finding meaning for differences in point estimates

Pros and Cons

Advantages of using PCIV:

- Estimate PAEs (& LATEs for an identified group of compliers) under less restrictive assumptions
- Estimate the distribution of effect heterogeneity
- Root mean squared error declines with more observations per-cluster
- PC first-stage provides insight regarding key assumptions underpinning IV estimation with grouped data

Pros and Cons

Advantages of using PCIV:

- Estimate PAEs (& LATEs for an identified group of compliers) under less restrictive assumptions
- Estimate the distribution of effect heterogeneity
- Root mean squared error declines with more observations per-cluster
- PC first-stage provides insight regarding key assumptions underpinning IV estimation with grouped data

Constraints to using PCIV:

- Need sufficiently large clusters
- Sacrifice efficiency for robustness

Pros and Cons

Advantages of using PCIV:

- Estimate PAEs (& LATEs for an identified group of compliers) under less restrictive assumptions
- Estimate the distribution of effect heterogeneity
- Root mean squared error declines with more observations per-cluster
- PC first-stage provides insight regarding key assumptions underpinning IV estimation with grouped data

Constraints to using PCIV:

- Need sufficiently large clusters
- Sacrifice efficiency for robustness

Model Identification Constraint:

- Need variation in each cluster for each cluster to be represented in the PAE without further assumptions

General Model

Consider a correlated random effects model as follows:

$$\begin{aligned} y_{ij} &= \mathbf{x}_{1ij}\mathbf{b}_i + \mathbf{x}_{2ij}\boldsymbol{\delta} + e_{ij}, \\ \mathbf{x}_{1ij} &= \mathbf{z}_{ij}\boldsymbol{\Gamma}_i + \mathbf{x}_{2ij}\boldsymbol{\eta} + u_{ij},\ i = 1, ..., N; j = 1, ..., T, \end{aligned} \tag{1}$$

- y_{ij} is a dependent variable
- $\mathbf{x}_{1ij}$ is a $1 \times K$ vector of plausibly endogenous covariates
  - Plausibly heterogeneous second-stage slopes $\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{d}_i$ such that $E(\mathbf{d}_i) = 0$
- $\mathbf{z}_{ij}$, a $1 \times L$ $(L \geq K)$ vector of instrumental variables
  - Plausibly heterogeneous first-stage slopes $\boldsymbol{\Gamma}_i = \boldsymbol{\gamma} + \mathbf{g}_i$ such that $E(\mathbf{g}_i) = 0$
- We allow for a $1 \times H$ vector of exogenous covariates, $\mathbf{x}_{2ij}$, with $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$, $H \times 1$ vectors of homogeneous slopes
  - May include a $1 \times T$ vector of time indicators s.t. $T < H$
- e_{ij} is an idiosyncratic error

General estimator: first stage (unbiased in finite samples)

1. Pre-treat the data to account for $\boldsymbol{\Gamma}_i$: Per-cluster, regress $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ij}$ on $\mathbf{z}_{ij}$, saving the residuals
2. To consistently estimate $\boldsymbol{\eta}$, regress the residuals $\tilde{\mathbf{x}}_{1ij}$ on $\tilde{\mathbf{x}}_{2ij}$ pooling over clusters to estimate $\hat{\boldsymbol{\eta}}$
3. Estimate $\boldsymbol{\Gamma}_i$ per-cluster by regressing $(\mathbf{x}_{1ij} - \mathbf{x_{2ij}}\hat{\boldsymbol{\eta}})$ on $\mathbf{z}_{ij}$ to obtain $\hat{\mathbf{X}}_{1i}$

1. Similar to estimating state-specific time trends and detrending

General estimator: first stage (unbiased in finite samples)

1. Pre-treat the data to account for $\mathbf{\Gamma}_i$: Per-cluster, regress $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ij}$ on $\mathbf{z}_{ij}$, saving the residuals
2. To consistently estimate $\boldsymbol{\eta}$, regress the residuals $\tilde{\mathbf{x}}_{1ij}$ on $\tilde{\mathbf{x}}_{2ij}$ pooling over clusters to estimate $\hat{\boldsymbol{\eta}}$
3. Estimate $\mathbf{\Gamma}_i$ per-cluster by regressing $(\mathbf{x}_{1ij} - \mathbf{x_{2ij}}\hat{\boldsymbol{\eta}})$ on $\mathbf{z}_{ij}$ to obtain $\hat{\mathbf{X}}_{1i}$

1. Similar to estimating state-specific time trends and detrending
2. Use cross-sectional variation to estimate homogeneous parameters

General estimator: first stage (unbiased in finite samples)

1. Pre-treat the data to account for $\boldsymbol{\Gamma}_i$: Per-cluster, regress $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ij}$ on $\mathbf{z}_{ij}$, saving the residuals
2. To consistently estimate $\boldsymbol{\eta}$, regress the residuals $\tilde{\mathbf{x}}_{1ij}$ on $\tilde{\mathbf{x}}_{2ij}$ pooling over clusters to estimate $\hat{\boldsymbol{\eta}}$
3. Estimate $\boldsymbol{\Gamma}_i$ per-cluster by regressing $(\mathbf{x}_{1ij} - \mathbf{x_{2ij}}\hat{\boldsymbol{\eta}})$ on $\mathbf{z}_{ij}$ to obtain $\hat{\mathbf{X}}_{1i}$

1. Similar to estimating state-specific time trends and detrending
2. Use cross-sectional variation to estimate homogeneous parameters
3. Use within variation (net of homogeneous parameters) to get the first stage by cluster (Wooldridge, 2010)

General estimator: second stage (rinse wash repeat)

1. For the second stage, we regress y_{ij} and $\mathbf{x}_{2ij}$ on $\hat{\mathbf{x}}_{1ij}$ per-cluster, obtaining the residuals $\dot{\mathbf{y}}_i$ and $\dot{\mathbf{x}}_{2i}$
2. Regressing the residuals $\dot{y}_{ij}$ on $\dot{\mathbf{x}}_{2ij}$ pooling over clusters allows us to eliminate $\mathbf{b}_i$ when we obtain $\hat{\boldsymbol{\delta}}$
3. The heterogeneous slopes $\hat{\mathbf{b}}_i = (\hat{\mathbf{X}}'_{1i}\hat{\mathbf{X}}_{1i})^{-1}\hat{\mathbf{X}}'_{1i}(\mathbf{y}_i - \mathbf{X}_{2i}\hat{\boldsymbol{\delta}})$ can be consistently estimated by regressing $(y_{ij} - \mathbf{x}_{2ij}\delta)$ on $\hat{\mathbf{x}}_{1ij}$ per cluster

General estimator: second stage (rinse wash repeat)

1. For the second stage, we regress y_{ij} and $\mathbf{x}_{2ij}$ on $\hat{\mathbf{x}}_{1ij}$ per-cluster, obtaining the residuals $\dot{\mathbf{y}}_i$ and $\dot{\mathbf{x}}_{2i}$
2. Regressing the residuals $\dot{y}_{ij}$ on $\dot{\mathbf{x}}_{2ij}$ pooling over clusters allows us to eliminate $\mathbf{b}_i$ when we obtain $\hat{\boldsymbol{\delta}}$
3. The heterogeneous slopes $\hat{\mathbf{b}}_i = (\hat{\mathbf{X}}'_{1i}\hat{\mathbf{X}}_{1i})^{-1}\hat{\mathbf{X}}'_{1i}(\mathbf{y}_i - \mathbf{X}_{2i}\hat{\boldsymbol{\delta}})$ can be consistently estimated by regressing $(y_{ij} - \mathbf{x}_{2ij}\delta)$ on $\hat{\mathbf{x}}_{1ij}$ per cluster
4. Averaging over $\hat{\mathbf{b}}_i$ obtains $\hat{\beta}_{PCIV} = \sum_{i=1}^{N} w_i \hat{\mathbf{b}}_i$

General estimator: second stage (rinse wash repeat)

1. For the second stage, we regress y_{ij} and $\mathbf{x}_{2ij}$ on $\hat{\mathbf{x}}_{1ij}$ per-cluster, obtaining the residuals $\dot{\mathbf{y}}_i$ and $\dot{\mathbf{x}}_{2i}$
2. Regressing the residuals $\dot{y}_{ij}$ on $\dot{\mathbf{x}}_{2ij}$ pooling over clusters allows us to eliminate $\mathbf{b}_i$ when we obtain $\hat{\boldsymbol{\delta}}$
3. The heterogeneous slopes $\hat{\mathbf{b}}_i = (\hat{\mathbf{X}}'_{1i}\hat{\mathbf{X}}_{1i})^{-1}\hat{\mathbf{X}}'_{1i}(\mathbf{y}_i - \mathbf{X}_{2i}\hat{\boldsymbol{\delta}})$ can be consistently estimated by regressing $(y_{ij} - \mathbf{x}_{2ij}\delta)$ on $\hat{\mathbf{x}}_{1ij}$ per cluster
4. Averaging over $\hat{\mathbf{b}}_i$ obtains $\hat{\beta}_{PCIV} = \sum_{i=1}^{N} w_i\hat{\mathbf{b}}_i$

pciv.ado Stata package coming...

Consistency of $\widehat{\beta}_{\mathrm{PCIV}}$

With $T \to \infty$ and $N \to \infty$,

$$\operatorname*{plim}_{T,\ N\to\infty} (\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}} - \boldsymbol{\beta}) = \mathrm{E}(\mathbf{d_i}) + \mathrm{E}[[\mathrm{E}_i(\mathbf{x'_{ij}}\mathbf{H_{z_i}}\mathbf{x_{ij}})]^{-1}\mathrm{E}_i(\mathbf{x'_{ij}}\mathbf{H_{z_i}}e_{ij})] \tag{2}$$

Thus, consistency of $\hat{\boldsymbol{\beta}}_{\boldsymbol{PCIV}}$ follows from the assumptions enumerated below:

(A1) $i.i.d.$ across i

(A2) $\mathrm{E}[e_{ij} \mid \mathbf{x}_{2i}, \mathbf{z}_i, \mathbf{d}_i] = 0$, $\mathrm{E}[u_{ij} \mid \mathbf{x}_{2i}, \mathbf{z}_i, \mathbf{g}_i] = 0$

(A3) $rank[\mathrm{E}_i(\mathbf{z}'_{ij}\mathbf{x}_{ij})] = K$, $rank[\mathrm{E}_i(\mathbf{z}'_{ij}\mathbf{z}_{ij})] = L$, and $\mathrm{E}[z'_{ij}z_{ij}e^2_{ij}]$ is positive definite

(A4) $\mathrm{E}[\|\mathbf{x_{2\mathit{ij}}}\|^2] < \infty$, $\mathrm{E}[\|\mathbf{z_{\mathit{ij}}}\|^2] < \infty$; $\mathrm{E}[\|\mathbf{z_{\mathit{ij}}}\|^4] < \infty$, and $\mathrm{E}[\|\mathbf{x_{2\mathit{ij}}}\|^4] < \infty$

(A5) $w_i = O_p(n^\epsilon)$ where $\sum_{i=1}^N w_i = 1$ and $\epsilon \leq -1$; $N^{3+2\epsilon}/T \to c$, where $0 < c < \infty$

Finite sample small T

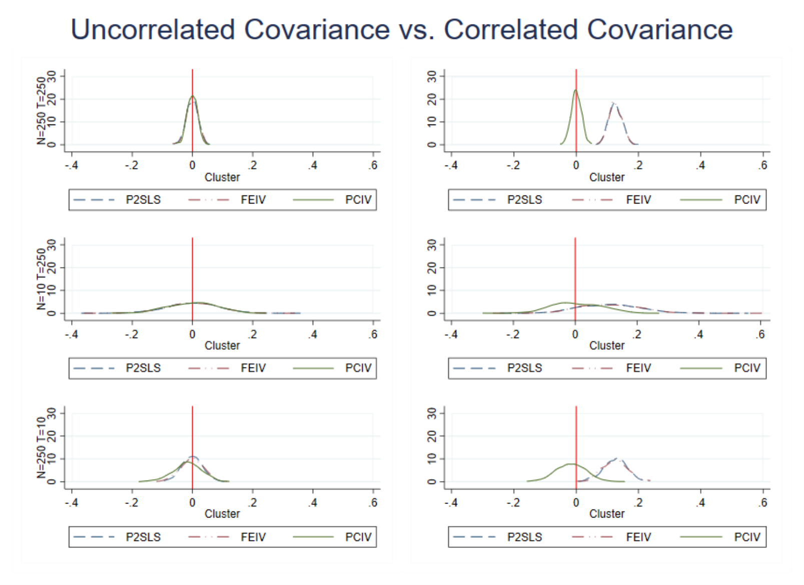
Finite sample small N

Inference

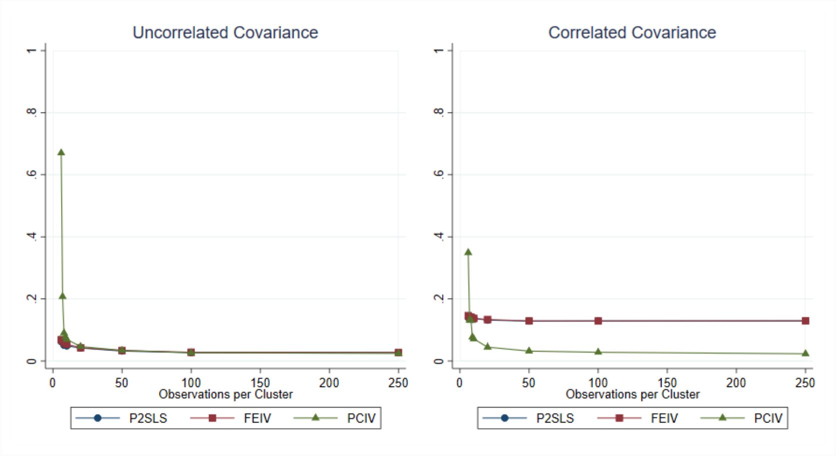
$$\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}_{PCIV} - \boldsymbol{\beta}) \tag{3}$$
$$= \mathrm{V}\left(\sum_{i=1}^{N} w_i \left[\hat{\mathbf{d}}_i + \left(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i}\right)^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\hat{e}_i\right]\right)$$
$$= \sum_{i=1}^{N} w_i^2\hat{\mathbf{d}}_i\hat{\mathbf{d}}_i' + \sum_{i=1}^{N} w_i^2\left(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i}\right)^{-1}\mathbf{X}_{1i}'\mathbf{P}_i\hat{\Omega}\mathbf{P}_i'\mathbf{X}_{1i}\left(\mathbf{X}_{1i}'\mathbf{P}_i\mathbf{X}_{1i}\right)^{-1}$$

where $\hat{\mathbf{d}}_{\mathbf{i}} = \hat{\mathbf{b}}_i - \hat{\boldsymbol{\beta}}_{PCIV}$ and $\hat{e}_i = y_i - \boldsymbol{X}_{1i}\hat{\boldsymbol{b}}_i - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\delta}}$. The standard errors from this estimator are robust to heteroskedasticity and arbitrary correlation in the error term within cluster

Simplified model: kernel density plots of estimation errors, $\hat{\beta}_1 - \beta_1$

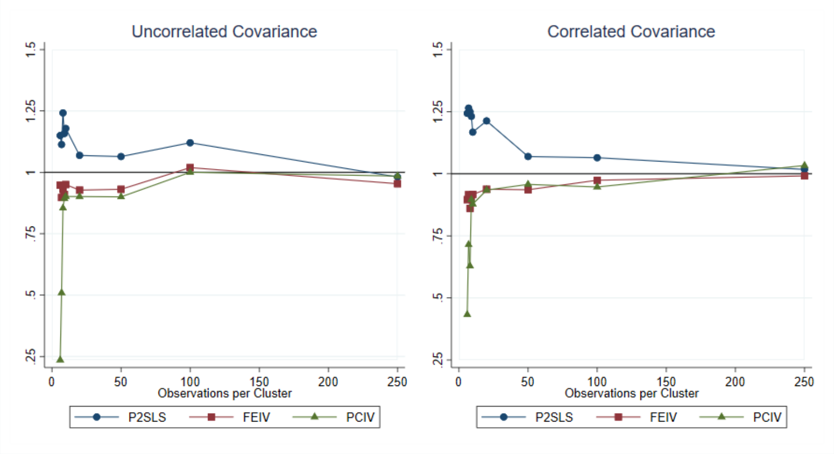


Root Mean Squared Errors across cluster size



DPG

Ratio of Mean SE by SD across cluster size



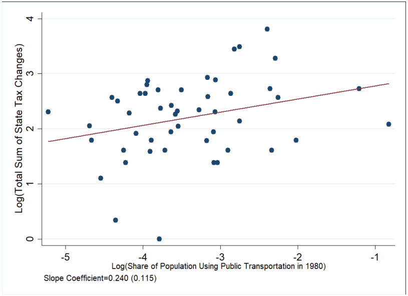
Note: Ratio of mean standard errors (SEs) divided by standard deviations (SDs) of the estimates

Bias Coverage rates Finite sample small N

PCIV in practice: Elasticity of demand for gasoline

- First-stage heterogeneity: sales and per-unit taxes vary by state
- Second-stage heterogeneity: infrastructure, population density, and local economies vary by state
- Correlation: States with more infrastructure raise taxes more

Figure: Relationship between tax changes and public transportation



Estimating equation

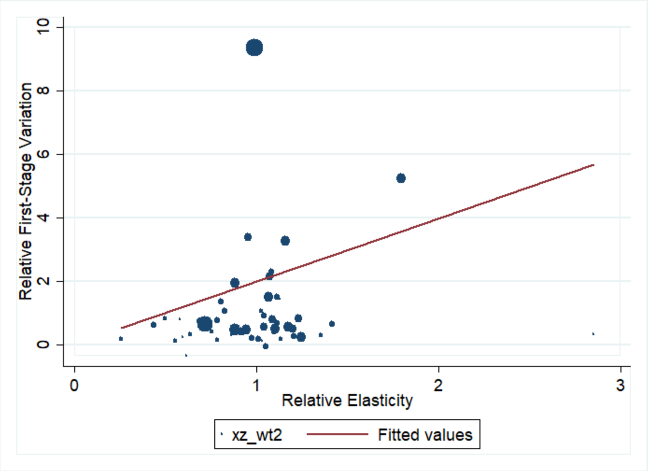
$$\begin{aligned} logsales_{ij} &= \alpha_{1i} + logprice_{ij} b_i + \mathbf{x}_{ij}\boldsymbol{\delta} + \tau_{1t} + \epsilon_{ij}, \\ logprice_{ij} &= \alpha_{2i} + logtaxes_{ij}\Gamma_i + \mathbf{x}_{ij}\boldsymbol{\eta}_2 + \tau_{2t} + u_{ij}. \end{aligned} \tag{4}$$

- Includes state and month-by-year fixed effects
- Includes population, income, unemployment, temperature, and rainfall exogenous covariates

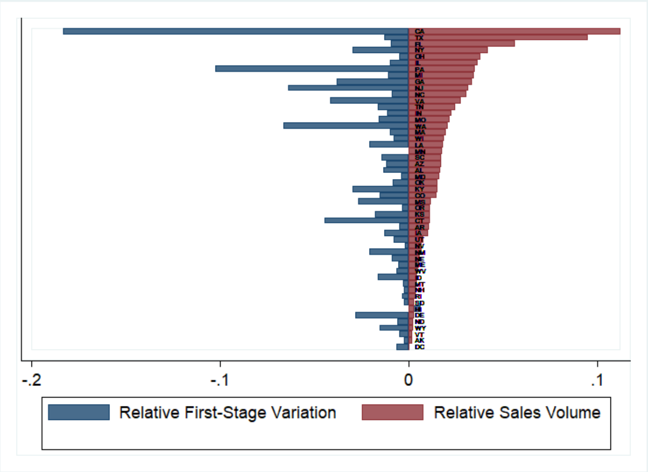
Summary of Results Using Three Estimation Methods

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-0.724	-0.929	-0.551	-0.463	-0.873	-0.555
	(0.193)	(0.415)	(0.227)	(0.154)	(0.394)	(0.240)
First-stage F-statistic	36.66	79.71	58.35	47.47	63.70	61.16
Controls	N	N	N	N	N	N
Log price	-0.736	-0.828	-0.543	-0.512	-0.760	-0.561
	(0.189)	(0.327)	(0.278)	(0.138)	(0.271)	(0.294)
First-stage F-statistic	36.58	80.92	58.71	46.83	60.26	59.93
Controls	Y	Y	Y	Y	Y	Y

Violation: correlated elasticities and first-stage variation



FEIV implicit weighting vs volume weights



LATE

Table: Estimated elasticities among states in which the instrument is strong (LATE)

	Without volume weights			Volume weighted		
	P2SLS	FEIV	PCIV	P2SLS	FEIV	PCIV
Log price	-0.659	-1.153	-0.521	-0.430	-0.945	-0.541
	(0.208)	(0.389)	(0.216)	(0.170)	(0.399)	(0.210)
First-stage F-statistic	43.24	80.92	55.91	98.21	64.51	58.35

Notes: Sample composed of all states with first-stage F-statistics above 10, excluding Hawaii, Indiana, Georgia, Michigan, and the District of Columbia. Regressions condition on time-by-month fixed effects. State-clustered standard errors appear in parentheses.

Conclusion

- This paper suggests Per-Cluster Instrumental Variable Approach to identify PAEs
- When the strength of the instrument is related to the heterogeneous effects, PCIV can consistently estimate Population Average Effects
- With access to large T data sets, PCIV strictly dominates FEIV
- Even without a large T it may be useful, and we advocate considering PC first stage
- It seems that gasoline consumption is more elastic than typically thought, though the confidence interval is wide

Consistency for Fixed Effects Estimators

- Wooldridge (2005) shows the conditions under which standard fixed effects estimators are consistent in estimating PAEs

$$\widehat{\beta_{FE}} = \beta + (\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{\ddot{x}'_{ij}\ddot{x}_{ij}})^{-1}[\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{\ddot{x}'_{ij}\ddot{x}_{ij}d_i} + \sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{\ddot{x}'_{ij}\ddot{e}_{ij}}] \tag{5}$$

- Assumptions for consistency for the simple case model is as follows.

$$E(\mathbf{e_{ij}}|\mathbf{x_{i1}}, ..., \mathbf{x_{ij}}, \mathbf{b_i}) = 0,\ t = 1, ..., T, \tag{6}$$

$$rankE(\sum_{t=1}^{T}\mathbf{\ddot{x}'_{ij}\ddot{x}_{ij}}) = K, \tag{7}$$

$$E[\mathbf{\ddot{x}'_{ij}\ddot{x}_{ij}d_i}] = 0, \tag{8}$$

where $\mathbf{\ddot{x}_{ij}} = \mathbf{x_{ij}} - T^{-1}\sum_{t=1}^{T}\mathbf{x_{ij}}$.

Per-Cluster Estimation with Exogenous Regressors

- However, equation 8 may not hold in important cases.
- Per-cluster estimation in this simple model needs only two steps (Bates et al. (2014))
  - First, estimate $\widehat{\mathbf{b_i}}$ for each cluster using OLS on only the within-cluster observations, such that

$$\widehat{\mathbf{b_i}} = \beta + \mathbf{d_i} + (\sum_{t=1}^{T} \mathbf{x_{ij}'} \mathbf{x_{ij}})^{-1} (\sum_{t=1}^{T} \mathbf{x_{ij}'} e_{ij}) \quad (9)$$

  - Second, average $\widehat{\mathbf{b_i}}$ over clusters.

$$\widehat{\beta_{PC}} = \beta + N^{-1} \sum_{i=1}^{N} \mathbf{d_i} + N^{-1} \sum_{i=1}^{N} [(\sum_{t=1}^{T} \mathbf{x_{ij}'} \mathbf{x_{ij}})^{-1} (\sum_{t=1}^{T} \mathbf{x_{ij}'} e_{ij})] \quad (10)$$

  - From the rank conditions, $E(\mathbf{d_i}) = 0$ by definition, and the strict exogeneity assumption from equation (2), per-cluster estimator is **unbiased**

Asymptotic unbiasedness with fixed N

With $T \to \infty$ and N fixed,

$$\begin{aligned} &\underset{T\to\infty}{plim}\ \widehat{\boldsymbol{\beta}_{\boldsymbol{PCIV}}} \\ &= \beta + N^{-1}\sum_{i=1}^{N}\mathbf{d_i} + N^{-1}\sum_{i=1}^{N}[(E(\mathbf{z'_{ij}}\mathbf{x_{ij}}))^{-1}E(\mathbf{z'_{ij}}e_{ij})]. \end{aligned} \tag{11}$$

$$E(\underset{T\to\infty}{plim}\ \widehat{\boldsymbol{\beta}_{\boldsymbol{PCIV}}}) = \boldsymbol{\beta} + E(\mathbf{d_i}) + E[[E_i(\mathbf{z'_{ij}}\mathbf{x_{ij}})]^{-1}E_i(\mathbf{z_{ij}}'e_{ij})]. \tag{12}$$

$$\therefore\ \widehat{\boldsymbol{\beta}_{\boldsymbol{PCIV}}} \text{ is asymptotically unbiased.}$$

Unlike FEIV, the PCIV estimator may provide an asymptoticly unbiased estimate of $\boldsymbol{\beta}$ even when $E[(\ddot{\mathbf{z}}'_{\mathbf{ij}}\ddot{\mathbf{x}}_{\mathbf{ij}})^{-1}\ddot{\mathbf{z}}'_{\mathbf{ij}}\ddot{\mathbf{x}}_{\mathbf{ij}}\mathbf{d_i}] \neq 0$

Assumption needed for consistency with T fixed

With fixed T,

$$\underset{N\to\infty}{plim}\ \widehat{\beta_{PCIV}} = \beta + \underset{N\to\infty}{plim}\ N^{-1}\sum_{i=1}^{N}\mathbf{d_i} + \underset{N\to\infty}{plim}\ N^{-1}\sum_{i=1}^{N}[(\sum_{t=1}^{T}\mathbf{z'_{ij}x_{ij}})^{-1}\sum_{t=1}^{T}\mathbf{z'_{ij}}e_{ij}]. \quad (13)$$

$$\underset{N\to\infty}{plim}\ \widehat{\beta_{PCIV}} = \beta + E[\mathbf{d_i}] + E[(\sum_{t=1}^{T}\mathbf{z'_{ij}x_{ij}})^{-1}\sum_{t=1}^{T}\mathbf{z_{ij}}'e_{ij}]. \quad (14)$$

In this case, in order for $\widehat{\beta_{PCIV}}$ to consistently estimate the PAE (β), we must assume

$$E[(\textstyle\sum_{t=1}^{T}\mathbf{z'_{ij}x_{ij}})^{-1}\sum_{t=1}^{T}\mathbf{z_{ij}}'e_{ij}] = 0$$

- Each estimated b_i is bound to manifest some degree of finite sample bias (Bound et al. (1995))

Simulation: Data Generating Process

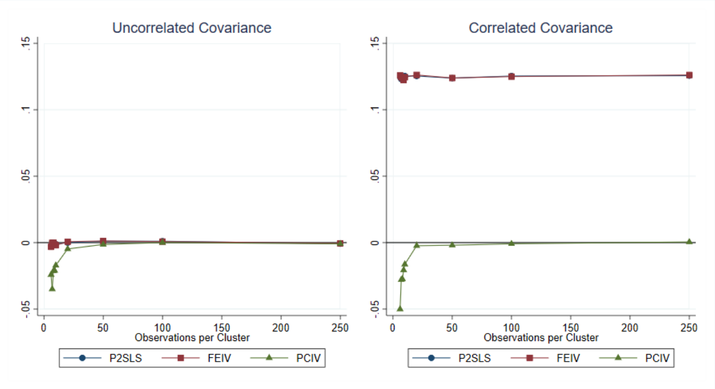
- We generate the data based on Equation 18 with 500 observations for each cluster
  - We generate two types of y_{ij} with $\beta_1 = 1$
    - Uncorrelated covariance assumption is violated with $y_{1,ij}$
    - Uncorrelated covariance assumption with $y_{2,ij}$

$$y_{1ij} = d_{0i} + (\beta_1 + d_{1i})x_{1ij} + o_{ij} + v_{ij}, \qquad v_{1ij} \sim N(0, \sigma_v). \tag{15}$$

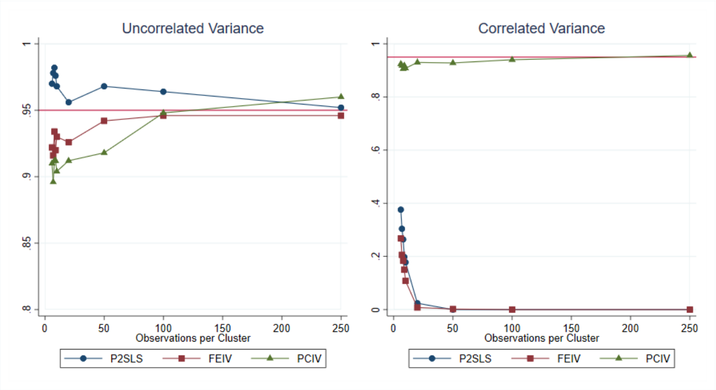
$$y_{2ij} = d_{0i} + (\beta_1 + d_{2i})x_{2ij} + o_{ij} + v_{2ij}, \qquad v_{2ij} \sim N(0, \sigma_v). \tag{16}$$

- x_{ij} is "observed" in the data and is a function $x_{2,it}$, z_{ij}, d_0, and d_2
- o_{ij} is exogenous but is "unobserved" in the data creating omitted variables bias
- $z_{ij} \sim N(0, \sigma_z)$. In order to violate the uncorrelated covariance assumption $\sigma_z = \exp(d_1)$
- d_0 and d_2 are each drawn from bivariate normal distribution and are allowed to be correlated with each other

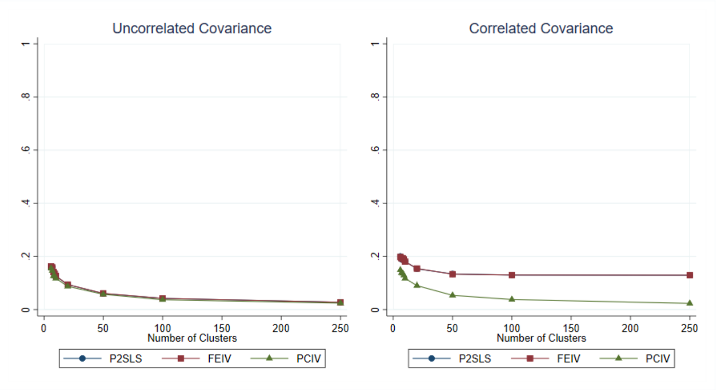
Simulation Results: Bias



Simulation Results: Coverage rate



Simulation Results: RMSE varying N of clusters



PC First Stage

- PC provides an unbiased estimate of the first stage.

$$\widehat{\gamma_{PC}} = \gamma + N^{-1} \sum_{i=1}^{N} \mathbf{d_i} + N^{-1} \sum_{i=1}^{N} [(\sum_{t=1}^{T} \mathbf{z_{ij}'} \mathbf{z_{ij}})^{-1} (\sum_{t=1}^{T} \mathbf{z_{ij}'} u_{ij})] \tag{17}$$

  - Allows identification of compliers
  - Tests monotonicity
  - Provides sample analogue to the key assumption behind FEIV

Weighting

Who is the population of interest?

- Panel data settings: Population from which the sample is drawn
  - However, panels are rarely random samples: NLSY and PSID both over-sample low income individuals and households
  - We can still recover PAEs using a weighted average in the last stage
- Grouped cross-sectional settings: Is the population individuals or groups?
  - If each groups comprise the population of interest with random sampling simple averaging is fine.
  - If we are interested in the individuals within groups, population weighted average in the last stage is needed to recover PAEs

Potential applications

Exogenous covariates

Consider the slightly richer model presented below:

$$y_{ij} = \mathbf{x_{1it}} \mathbf{b_i} + \mathbf{x_{2it}} \delta + e_{ij}, \; t = 1, ..., T, \tag{18}$$

- Year fixed effects provide one common and reasonable example of x_{2it}
- Including these in each regression could lead to the incidental parameters problem, and more generally cut into our degrees of freedom in the per-cluster regressions
- What to do?
  - Apply Frisch-Waugh-Lowell to residualize the data
  - Use residualized data for PCIV estimation

Potential applications

Mechanisms

Consider the following "multi-level model":

$$\begin{aligned} y_{ij} &= \mathbf{x_{ij}} \mathbf{b_i} + e_{ij}, \\ \mathbf{b_i} &= \beta + \mathbf{w_i}\gamma + \mathbf{d_i}, \end{aligned} \tag{19}$$

where $\mathbf{w_i} = (w_{1i}, ..., w_{Ji})$ is a vector of observable cluster-level components or mechanisms driving the heterogeneous effects

- If we assume $E[\mathbf{d_i} \mathbf{w_i}] = 0$ and $E[\mathbf{w_i} e_{ij}] = 0$, we can estimate these mechanisms using PCIV, allowing $E[\mathbf{x_{ij}} \mathbf{w_i}] \neq 0$.
- To estimate mechanisms:
  - Estimate cluster specific slopes $\widehat{\mathbf{b_{iPCIV}}}$ as before ignoring $\mathbf{w_i}$
  - In the second stage regress $\widehat{\mathbf{b_{iPCIV}}}$ on $\mathbf{w_i}$

Potential applications

Simulation Results Summary

- Under uncorrelated covariance, PCIV is noisier and seems to manifest additional finite sample bias with very small clusters. Still, with a cluster size of 20 all three estimators perform equally well

- In contrast, both FEIV and P2SLS are biased when the uncorrelated covariance assumption is violated. PCIV manifests less bias than the other two estimators across cluster sizes

- The ratio of mean SEs/SDs for PCIV is closest to 1, in both cases of uncorrelated and correlated covariance.

bib

Blundell, R., J. L. Horowitz, and M. Parey (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics 3*(1), 29–51.

Coglianese, J., L. W. Davis, L. Kilian, and J. H. Stock (2017). Anticipation, tax avoidance, and the price elasticity of gasoline demand. *Journal of Applied Econometrics 32*(1), 1–15.

Davis, L. W. and L. Kilian (2011). Estimating the effect of a gasoline tax on carbon emissions. *Journal of Applied Econometrics 26*(7), 1187–1214.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

Fernández-Val, I. and J. Lee (2013). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics 4*(3), 453–481.

Frondel, M., N. Ritter, and C. Vance (2012). Heterogeneity in the rebound effect: Further evidence for germany. *Energy*