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

Michael Bates[1], Jan Ditzen[2], and Seolah Kim[3]

UCR[1], Free University of Bozen-Bolzano[2], and CSULA[3]

STATA Conference 2025

# Motivating Question

**How do we best measure price elasticity of gasoline demand in the United States?**

The goal of this paper is to identify the average effect using gasoline tax in estimating the price elasticity.

# Motivating Question

**How do we best measure price elasticity of gasoline demand in the United States?**

The goal of this paper is to identify the average effect using gasoline tax in estimating the price elasticity.

Two potential challenges:

- Price and quantity are endogenous.
- Each state would have a different price elasticity of gasoline demand.

# What is Population Average Effect (PAE)?

- PAE is the average causal relationship between two variables over an entire population of interest.
- This is different from Local Average Treatment Effects (LATEs), which estimate the effect of compliers only.
- Heckman and Vytlacil (1998) coined the terminology of Correlated Random Coefficient (CRC) models for this environment: $y_i = \alpha_i + x_i(\beta + d_i) + e_i$.

# Estimating Population Average Effects

- **Panel or grouped cross-sectional data** allow estimating population average effects (PAEs) without imposing much structure.

- Murtazashvili and Wooldridge (2008) estimate PAEs with **endogenous regressors** using FEIV approaches.

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

- An important assumption is that

  " $\ddot{\mathbf{z}}_{\mathbf{ij}}' \ddot{\mathbf{x}}_{ij}$ is uncorrelated with $\mathbf{d}_{\mathbf{i}}$, "

  which means that the **strength of the instrument** should be uncorrelated with the **heterogeneous effects**.

# Model

Our Correlated Random Coefficient model is as follows:

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

where $y_{ij}$ is a dependent variable and $e_{ij}$ is an idiosyncratic error.

- The $1 \times K$ vector of endogenous variables, $\mathbf{x_{1ij}}$, includes 1; $\mathbf{z}_{ij}$, a $1 \times L$ $(L \geq K)$ vector of instrumental variables; and $1 \times H$ vector of exogenous covariates, $\mathbf{x}_{2ij}$.
- A key feature of the model in cluster-specific slopes,

$$\mathbf{b_i} = \beta + \mathbf{d_i}, \text{ where } E(\mathbf{d_i}) = 0$$
$$\gamma_i = \gamma + \mathbf{g}_i, \text{ where } E(\mathbf{g}_i) = 0$$

These indicate the heterogeneous effects that vary by cluster.

# Proposed estimator: Per-Cluster Instrumental Variables

From the first-stage equation ($\mathbf{x}_{1ij} = \mathbf{z}_{ij}\boldsymbol{\gamma}_i + \mathbf{x}_{2ij}\boldsymbol{\eta} + u_{ij}$),

**STEP 1:**
For each cluster, regress $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ij}$ on $\mathbf{z}_{ij}$ separately, then obtain the residuals denoted as $\tilde{\mathbf{x}}_{1ij}$ and $\tilde{\mathbf{x}}_{2ij}$.

**STEP 2:**
Estimate the equation, $\tilde{\mathbf{x}}_{1ij} = \tilde{\mathbf{x}}_{2ij}\boldsymbol{\eta} + \epsilon_{ij}$ using the pooled sample, and obtain $\hat{\boldsymbol{\eta}}$.

# Proposed estimator: PCIV Approach (Cont'd)

**STEP 3:**

With the estimated $\hat{\boldsymbol{\eta}}$, estimate the following equation to estimate $\gamma_i$ per cluster:

$$(\mathbf{x}_{1ij} - \mathbf{x_{2ij}}\hat{\boldsymbol{\eta}}) = \mathbf{z}_{ij}\gamma_i + \xi_{ij}$$

**STEP 4:**

Using the fitted values from the first-stage regression ($\hat{\mathbf{x}}_{1ij}$), repeat the procedure for the second-stage equation to get $\hat{\mathbf{b}}_{i,PCIV}$.

**STEP 5:**

We can get the PAE estimate, $\hat{\boldsymbol{\beta}}_{PCIV} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{b}}_{i,PCIV}$. We can also get a weighted estimator by $\hat{\boldsymbol{\beta}}_{PCIV} = \sum_{i=1}^{N} w_i\hat{\mathbf{b}}_i$.

# Per-Cluster Instrumental Variable Approach (PCIV)

Advantages of using PCIV:

- Estimate PAEs under less restrictive assumptions.
  $\rightarrow$ The strength of instruments can be correlated with heterogeneous effects.
- Performs well with more observations for each cluster.

Constraints to using PCIV:

- Need sufficiently large clusters.

# Syntax

pciv depvar [indepvars] (endovars = instvars), cluster(varname) options

- indepvars: common parameters
- endovars: list of endogenous variables
- instvars: list of instrumental variables
- cluster: variable to define individual slopes
- options: wt(varname), first, rf

# Syntax

pciv depvar [indepvars] (endovars = instvars), cluster(varname) options

- indepvars: common parameters
- endovars: list of endogenous variables
- instvars: list of instrumental variables
- cluster: variable to define individual slopes
- options: wt(varname), first, rf

$\rightarrow$ It can handle unbalanced panels.

# Examples

- Display first-stage and reduced-form regression results:
  pciv logvolume dat (logprice = logtax), cluster(statefip) first rf
- With weights:
  pciv logvolume dat (logprice = logtax), cluster(statefip) wt(wt)
- With multiple endogenous variables and instrumental variables:
  pciv logvolume (logprice flogprice = logtax flogtax), cluster(statefip)
- Without endogenous variables:
  pciv logvolume dat ( = logtax), cluster(statefip)

# Post-Estimation Commands

- N_g, N, Tmin, Tmax, vce
- predict: fitted values, residuals, rform, instruments
- Cluster-level coefficients
  - reduced-form, second-stage estimates
- Future tasks: $R^2$, F-Stats, allowing factor variables, adding weak IV test, Hansen's J-test, etc.

# Example Results

```
. pciv logvolume tm1-tm359 dat (logprice = logtax), cluster(statefip)

Second-Stage Regression Results

Observations: 18360
Groups: 51
T (min): 360
T (max): 360
-------------------------------------------------------
Variable      |       Coef. | Std. Err. |       t | P>|t|
-------------------------------------------------------
logprice      |      -0.536 |     0.203 |   -2.65 | 0.011
tm1           |      -0.584 |     0.131 |   -4.46 | 0.000
tm2           |      -0.565 |     0.121 |   -4.67 | 0.000
tm3           |      -0.470 |     0.134 |   -3.50 | 0.001
tm4           |      -0.445 |     0.138 |   -3.23 | 0.002
tm5           |      -0.405 |     0.147 |   -2.76 | 0.008
tm6           |      -0.341 |     0.171 |   -1.99 | 0.052
tm7           |      -0.391 |     0.165 |   -2.37 | 0.021
tm8           |      -0.362 |     0.156 |   -2.32 | 0.025
tm9           |      -0.426 |     0.141 |   -3.01 | 0.004
tm10          |      -0.459 |     0.138 |   -3.33 | 0.002
```

# Example Results (Cont'd)

Figure: Summary of Group Coefficients

```
. estat groupcoeffs
```

**Group-Specific Coefficients (Reduced-Form)**

| Variable | Mean | Std. Dev. | Min | Max | N |
|---|---|---|---|---|---|
| logtax | -0.240 | 0.529 | -1.251 | 2.466 | 51 |

**Group-Specific Coefficients (Second-Stage)**

| Variable | Mean | Std. Dev. | Min | Max | N |
|---|---|---|---|---|---|
| logprice | -0.536 | 0.205 | -1.536 | -0.139 | 51 |

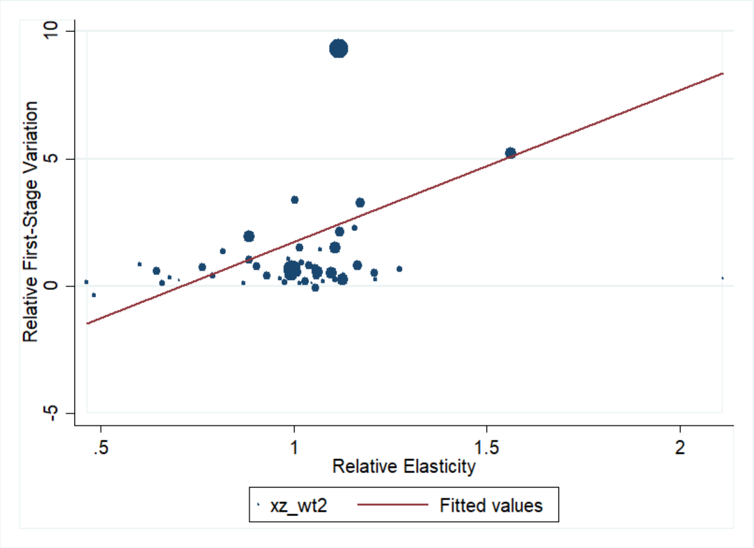
# Estimating the Price Elasticity Demand for Gasoline

From Bates & Kim (2024, JAE),

$$\log sales_{ij} = \alpha_{1i} + \log price_{ij}\mathbf{b}_i + \mathbf{x}_{ij}\boldsymbol{\delta} + \epsilon_{ij},$$
$$\log price_{ij} = \alpha_{2i} + \log taxes_{ij}\gamma_i + \mathbf{x}_{ij}\boldsymbol{\eta} + u_{ij}.$$

- We are primarily interested in the population average price elasticity of gasoline demand, $E[\mathbf{b}_i] = \boldsymbol{\beta}$.
- We use the log of taxes as instruments for the potentially endogenous log of prices. We allow for possible heterogeneity in tax pass-through rates, as denoted by $\gamma_i$.

# Heterogeneous Elasticities and First-Stage Variation



# Summary of Results Using Three Estimation Methods

Table: 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 |

*Notes*: The sample consists of 18,360 state-by-month observations. First-stage F-statistics for P2SLS and FEIV are obtained from the regression of each endogenous regressor on the exogenous regressors and the instruments. The calculation of the first-stage F-statistics for the PCIV was done using Hotelling's T-squared test. 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 correlated to the heterogeneous effects, PCIV can consistently estimate Population Average Effects.
- The development of the STATA package is underway. Any feedback is welcome!

**Thank you!**