

*Selection of Heterogeneous Instruments in Partially Linear  
Fixed Effects Panel Regression*

*or the clickbait version:*

*Can You Select Heterogeneous Instruments in a Partially Linear Fixed  
Effects Panel Regression? A Machine Learning Approach*

By: Eric Penner

September 17, 2018

## Introduction pt.1

### Basic Elements

- ▶ Linear in parameters fixed effects panel model
- ▶ Subset of regressors are endogenous
- ▶ Relevant instruments vary by cross section
- ▶ Relevant instruments are unknown subset of larger collection

## Introduction pt.2

### Applications / Motivation

- ▶ Growth and Foreign Aid Models: Burnside and Dollar (2000) AER
- ▶ Sensitivity to Set of Instruments: Leon-Gonzalez et. al. (2015) JoMa
- ▶ Making model selection more explicit while battling overfitting
- ▶ Sufficient spatial distance separating cross sections
- ▶ Little to no domain knowledge / replacing ad hoc procedures

*Initial Model*

$$Y_{jt} = \beta_0 + Z'_{1jt}\beta_1 + Z'_{2jt}\beta_2 + e_j + \varepsilon_{jt} \quad (1)$$

$$Z_{1jdt} = \alpha_{jd0} + Z'_{2jt}\alpha_{1d} + W'_{jdt}\alpha_{2jd} + V_{jdt} \quad (2d)$$

$$E(\varepsilon_{jt}|Z_{1jt}, Z_{2jt}) = E(\varepsilon_{jt}|Z_{1jt}) \neq 0 \quad (3)$$

$$E(V_{jdt}|Z_{2jt}, W_{jt}) = 0 \quad (4)$$

- ▶  $Y_{jt}$  is a scalar random variable
- ▶  $Z_{1jt} \in \mathbb{R}^{p_1}$ ,  $Z_{2jt} \in \mathbb{R}^{p_2}$  are endogenous and exogenous variables respectively
- ▶  $e_j$  is a fixed effect,  $\alpha_{jd0}$  is a combined constant and fixed effect
- ▶  $W_t \in \mathbb{R}^w$  a vector of instrumental variables.
- ▶  $W_{jdt} \in \mathbb{R}^{w_{jd}}$  unknown subvector of  $W_t \in \mathbb{R}^w$ .

## Primary Assumptions

- i.) Exclusion Restriction:  $E(\varepsilon_{jt}|Z_{1jt}) = E(\varepsilon_{jt}|Z_{2jt}, W_t, V_{jt}) = E(\varepsilon_{jt}|V_{jt})$
- ii.) Control Function:  $\varepsilon_{jt} = f_j(V_{jt}) + u_{jt}$
- iii.) Orthogonality:  $E[u_{jt}|Z_{1jt}, Z_{2jt}, W_t] = 0$
- iv.) Additivity:  $f_j(V_{jt}) = \sum_d f_{jd}(V_{jdt})$
- v.) Panel Secondary Equation:  $W'_{jdt}\alpha_{2jd} = W'_t M_{jd}\alpha_{2d}$  where  $M_{jd} = \text{diag}(m_{jd})$  and
 
$$m_{jd} = [1\{W_{1t} \in W_{jdt}\} \quad 1\{W_{2t} \in W_{jdt}\} \quad \cdots \quad 1\{W_{wt} \in W_{jdt}\}]'$$

Consequently,

$$E(\varepsilon_{jt}|Z_{2jt}, W_t, V_{jt}) = E(\varepsilon_{jt}|V_{jt}) = E(f_j(V_{jt}) + u_{jt}|V_{jt}) = f_j(V_{jt}) = \sum_{d=1}^{p_1} f_{jd}(V_{jdt})$$

*First Differenced Model*

Let  $\Delta$  be the first difference in operator, and assume  $\{Y_{jt}, Z_{1jt}, Z_{2jt}, W_t\}_{t=1}^n$  is i.i.d

$$\Delta Y_{jt} = \Delta Z'_{1jt} \beta_1 + \Delta Z'_{2jt} \beta_2 + \sum_{d=1}^{p_1} \Delta f_{jd}(V_{jdt}) + \Delta u_{jt} \quad (5)$$

$$Z_{1jdt} = \alpha_{jd0} + Z'_{2jt} \alpha_{1d} + W'_t M_{jd} \alpha_{2d} + V_{jdt} \quad (2d)$$

$$E(\Delta u_{jt} | V_{jt}, V_{j,t-1}) = 0 \quad (7)$$

$$E(V_{jdt} | Z_{2jt}, W_{jt}) = 0 \quad (8)$$

Residuals on residuals regression: Manzan and Zerom (2005) SaPL

## Projections and Notation

Density Ratio

$$\phi_{jt} = \frac{\prod_{d=1}^{p_1} \rho(V_{jdt}, V_{jd(t-1)})}{\rho(V_{jt}, V_{j(t-1)})}$$

Conditional Expectations, for  $k \in \{1, 2\}$

$$H_{jd}(\Delta Z_{kjt}) = E[\phi_{jt} \Delta Z_{kjt} | V_{jdt}, V_{jd(t-1)}]$$

$$H_{jd}(\Delta Y_{jt}) = E[\phi_{jt} \Delta Y_{jt} | V_{jdt}, V_{jd(t-1)}]$$

$$H_j(\Delta Z_{kjt}) = \sum_{d=1}^{p_1} H_{jd}(\Delta Z_{kjt})$$

$$H_j(\Delta Y_{jt}) = \sum_{d=1}^{p_1} H_{jd}(\Delta Y_{jt})$$

Vectors

$$H(\Delta Y_t) = [H_1(\Delta Y_{1t}) \ H_2(\Delta Y_{2t}) \ \cdots \ H_q(\Delta Y_{qt})]'$$

$$H(\Delta Z_{kt}) = [H_1(\Delta Z_{k1t}) \ H_2(\Delta Z_{k2t}) \ \cdots \ H_q(\Delta Z_{kqt})]'$$

$$\Delta Y_t = [\Delta Y_{1t} \ \Delta Y_{2t} \ \cdots \ \Delta Y_{qt}]'$$

$$\Delta Z_{kt} = [\Delta Z_{k1t} \ \Delta Z_{k2t} \ \cdots \ \Delta Z_{kqt}]'$$

# Identification Lemma

## Lemma: Identification

Let  $\phi_t = \text{diag}(\{\phi_{jt}\}_{j=1}^q)$ , and

$$\Delta Z_t = [\Delta Z'_{1t} \quad \Delta Z'_{2t}]' \quad H(\Delta Z_t) = [H(\Delta Z_{1t})' \quad H(\Delta Z_{2t})']'$$

Then if,

i.) for all  $d \in \{1, 2, \dots, p_1\}$ ,  $j \in \{1, 2, \dots, q\}$ , and  $t \in \{2, \dots, T\}$

$$E[f_{jd}(V_{jdt})] = E[f_{jd}(V_{jd(t-1)})]$$

ii.)  $E([\Delta Z_t - H(\Delta Z_t)]' \phi_t [\Delta Z_t - H(\Delta Z_t)])$  is positive semi definite,

Then  $\beta_1$  and  $\beta_2$  are identified, in particular,

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = E([\Delta Z_t - H(\Delta Z_t)]' \phi_t [\Delta Z_t - H(\Delta Z_t)])^{-1} E([\Delta Z_t - H(\Delta Z_t)]' \phi_t [\Delta Y_t - H(\Delta Y_t)])$$



## Secondary Equation Estimation pt.1

Must Estimate  $\hat{V}_{jdt}$  as a first step

- ▶  $W_{jdt} \in \mathbb{R}^{w_{jd}}$  unknown subvector of  $W_t \in \mathbb{R}^w$ .
- ▶ Subset selection and coefficient estimation with lasso

Why Subset / Model Selection?

- ▶ Overfitting  $E(Z_{1jdt}|Z_{2jt}, W_t)$
- ▶ If  $T \ll w$

Lasso: Is the treatment worse than the original problem?

- ▶ Underfitting  $E(Z_{1jdt}|Z_{2jt}, W_t)$
- ▶ How to estimate  $\alpha_{2d}$  when  $M_{jd}$  is also unknown?

## *Model Selection and Overfitting*



## Sample Splitting

- ▶ Let  $n_p \in \mathbb{N}$  be the number of partitions to be generated
- ▶ Generate Partition  $\{\mathcal{I}_g\}_{g=1}^{n_p}$  of  $\{1, 2, \dots, T\}$  i.e.

$$\mathcal{I}_g \subset \{i\}_{i=1}^T \quad \mathcal{I}_g \cap \mathcal{I}_{g'} = \emptyset \quad \text{and} \quad \bigcup_{g=1}^{n_p} \mathcal{I}_g = \{i\}_{i=1}^T$$

where each partition has approx same cardinality

- ▶ Model Selection and Projection Estimation with Training Set:  $\mathcal{I}_g^c$
- ▶ Estimation of  $[\hat{\beta}'_{g1} \quad \hat{\beta}'_{g2}]'$  with Testing Set:  $\mathcal{I}_g$
- ▶ Final Estimator:

$$[\hat{\beta}'_1 \quad \hat{\beta}'_2]' = n_p^{-1} \sum_{g=1}^{n_p} [\hat{\beta}'_{g1} \quad \hat{\beta}'_{g2}]'$$

*Secondary Equation Estimation pt.2*

First Differenced Model

$$\Delta Z_{1jdt} = \Delta Z'_{2jt} \alpha_{1d} + \Delta W'_t M_{jd} \alpha_{2d} + \Delta V_{jdt}$$

Lasso Estimator with Training Set  $\mathcal{I}_g^c$

$$(\hat{\alpha}_{1d}, \hat{\alpha}_{2jd}) = \arg \min \sum_{j=1}^q \sum_{i \in \mathcal{I}_g^c} (\Delta Z_{1jdi} - \Delta Z'_{2ji} \alpha_1 - \Delta W'_i \alpha_{2j})^2 \quad \text{s.t.} \quad \sum_{l=1}^w |\alpha_{2j,l}| \leq \lambda$$

for all  $1 \leq j \leq q$ .

## Secondary Equation Estimation pt.3

Next define for some  $\varepsilon > 0$

$$\tilde{\alpha}_{2d,l} = \frac{\sum_{j=1}^q \hat{\alpha}_{2jd,l} \mathbf{1}\{|\hat{\alpha}_{2jd,l}| > \varepsilon\}}{\sum_{j=1}^q \mathbf{1}\{|\hat{\alpha}_{2jd,l}| > \varepsilon\} + \mathbf{1}\{\sum_{j=1}^q \mathbf{1}\{|\hat{\alpha}_{2jd,l}| > \varepsilon\} = 0\}}$$

Now let

$$\tilde{\alpha}_{2d} = [\tilde{\alpha}_{2d,1} \quad \tilde{\alpha}_{2d,2} \quad \cdots \quad \tilde{\alpha}_{2d,w}]'$$

and

$$\tilde{m}_{jd} = [\mathbf{1}\{|\hat{\alpha}_{2jd,1}| > \varepsilon\} \quad \mathbf{1}\{|\hat{\alpha}_{2jd,2}| > \varepsilon\} \quad \cdots \quad \mathbf{1}\{|\hat{\alpha}_{2jd,w}| > \varepsilon\}]'$$

so that, for all  $t \in \{1, 2, \dots, T\}$

$$\hat{V}_{jdt} = Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{1d} - W'_t \tilde{M}_{jd} \tilde{\alpha}_{2d} - \#(\mathcal{I}_g^c)^{-1} \sum_{i \in \mathcal{I}_g^c} (Z_{1jdi} - Z'_{2ji}\hat{\alpha}_{1d} - W'_i \tilde{M}_{jd} \tilde{\alpha}_{2d})$$

where  $\tilde{M}_{jd} = \text{diag}(\tilde{m}_{jd})$

## Density Ratio Estimation

Density Estimation: for all  $j \in \{1, 2, \dots, q\}$ ,  $d \in \{1, 2, \dots, p_1\}$ , and  $t \in \{2, \dots, T\}$

$$\hat{\rho}(\hat{V}_{jdt}, \hat{V}_{jd(t-1)}) = (n(t)h_1^2)^{-1} \sum_{i \in \mathcal{I}_g^c, i \neq t} k\left(\frac{\hat{V}_{jdi} - \hat{V}_{jdt}}{h_1}\right) k\left(\frac{\hat{V}_{jd(i-1)} - \hat{V}_{jd(t-1)}}{h_2}\right)$$

$$\hat{\rho}(\hat{V}_{jt}, \hat{V}_{j(t-1)}) = (n(t)h_2^{2p_1})^{-1} \sum_{i \in \mathcal{I}_g^c, i \neq t} \prod_{d=1}^{p_1} k\left(\frac{\hat{V}_{jdi} - \hat{V}_{jdt}}{h_2}\right) k\left(\frac{\hat{V}_{jd(i-1)} - \hat{V}_{jd(t-1)}}{h_2}\right)$$

where

$$n(t) = \begin{cases} \#(\mathcal{I}_g^c) & \text{if } t \in \mathcal{I}_g \\ \#(\mathcal{I}_g^c) - 1 & \text{if } t \in \mathcal{I}_g^c \end{cases}$$

Density Ratio Construction:

$$\hat{\phi}_{jt} = \frac{\prod_{d=1}^{p_1} \hat{\rho}(\hat{V}_{jdt}, \hat{V}_{j(t-1),d})}{\hat{\rho}(\hat{V}_{jt}, \hat{V}_{j(t-1)})} \quad \hat{\theta}_{jdt} = \frac{\prod_{l \neq d}^{p_1} \hat{\rho}(\hat{V}_{jlt}, \hat{V}_{jl(t-1)})}{\hat{\rho}(\hat{V}_{jdt}, \hat{V}_{j(t-1)})}$$

## H Function Estimation

H Function Estimation: for each  $a \in \{1, 2\}$ ,  $j \in \{1, 2, \dots, q\}$ ,  
 $d \in \{1, 2, \dots, p_1\}$ ,  $\ell \in \{1, 2, \dots, p_a\}$ , and  $k \in \mathcal{I}_g$

$$\hat{H}_{jd}(\Delta Z_{aj\ell k}) = [n_{\mathcal{I}_g^c} h_3]^{-1} \sum_{i \in \mathcal{I}_g^c} k \left( \frac{\hat{V}_{jdi} - \hat{V}_{jdk}}{h_3} \right) k \left( \frac{\hat{V}_{jd(i-1)} - \hat{V}_{jd(k-1)}}{h_3} \right) \hat{\theta}_{jdi} \Delta Z_{aj\ell i}$$

and for each  $j \in \{1, 2, \dots, q\}$ ,  $d \in \{1, 2, \dots, p_1\}$ , and  $k \in \mathcal{I}_g$

$$\hat{H}_{jd}(\Delta Y_{jk}) = [n_{\mathcal{I}_g^c} h_3]^{-1} \sum_{i \in \mathcal{I}_g^c} k \left( \frac{\hat{V}_{jdi} - \hat{V}_{jdk}}{h_3} \right) k \left( \frac{\hat{V}_{jd(i-1)} - \hat{V}_{jd(k-1)}}{h_3} \right) \hat{\theta}_{jdi} \Delta Y_{ji}$$

then construct

$$\hat{H}_j(\Delta Z_{aj\ell k}) = \sum_{d=1}^{p_1} \hat{H}_{jd}(\Delta Z_{aj\ell k}) \quad \hat{H}_j(\Delta Y_{jk}) = \sum_{d=1}^{p_1} \hat{H}_{jd}(\Delta Y_{jk})$$

the rest is book keeping.

## Final Estimator for Testing Set $\mathcal{I}_g$

Let  $\hat{\phi} = \text{diag}(\{\hat{\phi}_k\}_{k \in \mathcal{I}_g})$

$$\begin{bmatrix} \hat{\beta}_{g1} \\ \hat{\beta}_{g2} \end{bmatrix} = \left( [\Delta Z - \hat{H}(\Delta Z)]' \hat{\phi} [\Delta Z - \hat{H}(\Delta Z)] \right)^{-1} \left( [\Delta Z - \hat{H}(\Delta Z)]' \hat{\phi} [\Delta Y - \hat{H}(\Delta Y)] \right)$$

Comparing the sampling distribution of  $\hat{\beta}_1$  and  $\hat{\beta}_2$ .

- ▶ Oracle v.s. Known Subset v.s. Unknown Subset v.s. Lasso
- ▶ Varying total number of instruments available  $w$
- ▶ Varying number of cross sections  $q$
- ▶ Varying number of time periods  $T$



## Equivalences and Covariances

Let

- ▶  $n_{tp} \equiv T$  be the total number of time periods.
- ▶  $n_{end} \equiv p_1$  be the number of endogenous regressors
- ▶  $n_{exo} \equiv p_2$  be the number of exogenous regressors
- ▶  $n_{tinst} \equiv w$  be the total number of available instruments
- ▶  $n_{cinst} \equiv w_j$  the number of instruments relevant to each crossection

$$\rho_{er} = \begin{bmatrix} \rho_{er,1} & \rho_{er,2} & \cdots & \rho_{er,n_{end}} \end{bmatrix}$$

$$\rho_{inst} = \begin{bmatrix} \rho_{inst,1} & \rho_{inst,2} & \cdots & \rho_{inst,n_{inst}-1} \end{bmatrix}$$

$$\rho_{ex} = \begin{bmatrix} \rho_{ex,1} & \rho_{ex,2} & \cdots & \rho_{ex,n_{ex}-1} \end{bmatrix}$$

be vectors of covariances.

## Error Covariance Matrix

For each cross section

$$V_{er} = \begin{bmatrix} 1 & \rho_{er,1} & \rho_{er,2} & \cdots & \rho_{er,n_{end}} \\ \rho_{er,1} & 1 & \rho_{er,1} & \cdots & \rho_{er,n_{end}-1} \\ \rho_{er,2} & \rho_{er,1} & 1 & \cdots & \rho_{er,n_{end}-2} \\ \vdots & & & \ddots & \\ \rho_{er,n_{end}} & \rho_{er,n_{end}-1} & \rho_{er,n_{end}-2} & \cdots & 1 \end{bmatrix}$$

For all cross sections

$$CV_{er} = \begin{bmatrix} V_{er} & \mathbf{0}_{(n_{end}+1 \times n_{end}+1)} & \cdots & \mathbf{0}_{(n_{end}+1 \times n_{end}+1)} \\ \mathbf{0}_{(n_{end}+1 \times n_{end}+1)} & V_{er} & \cdots & \mathbf{0}_{(n_{end}+1 \times n_{end}+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{(n_{end}+1 \times n_{end}+1)} & \mathbf{0}_{(n_{end}+1 \times n_{end}+1)} & \cdots & V_{er} \end{bmatrix}$$

## Exogenous Variable Covariance Matrix

For each cross section

$$V_{ex} = \begin{bmatrix} 1 & \rho_{ex,1} & \rho_{ex,2} & \cdots & \rho_{ex,n_{ex}-1} \\ \rho_{ex,1} & 1 & \rho_{ex,1} & \cdots & \rho_{ex,n_{ex}-2} \\ \rho_{ex,2} & \rho_{ex,1} & 1 & \cdots & \rho_{ex,n_{ex}-3} \\ \vdots & & & \ddots & \\ \rho_{ex,n_{ex}-1} & \rho_{ex,n_{ex}-2} & \rho_{ex,n_{ex}-2} & \cdots & 1 \end{bmatrix}$$

For all cross sections

$$CV_{ex} = \begin{bmatrix} V_{ex} & \mathbf{0}_{(n_{ex} \times n_{ex})} & \cdots & \mathbf{0}_{(n_{ex} \times n_{ex})} \\ \mathbf{0}_{(n_{ex} \times n_{ex})} & V_{ex} & \cdots & \mathbf{0}_{(n_{ex} \times n_{ex})} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{(n_{ex} \times n_{ex})} & \mathbf{0}_{(n_{ex} \times n_{ex})} & \cdots & V_{ex} \end{bmatrix}$$

## Common Instrument Covariance Matrix

$$V_{inst} = \begin{bmatrix} 1 & \rho_{inst,1} & \rho_{inst,2} & \cdots & \rho_{inst,n_{tinst}-1} \\ \rho_{inst,1} & 1 & \rho_{inst,1} & \cdots & \rho_{inst,n_{tinst}-2} \\ \rho_{inst,2} & \rho_{tinst,1} & 1 & \cdots & \rho_{inst,n_{tinst}-3} \\ \vdots & & & \ddots & \\ \rho_{inst,n_{tinst}-1} & \rho_{inst,n_{tinst}-2} & \rho_{inst,n_{tinst}-3} & \cdots & 1 \end{bmatrix}$$

## Exogenous Variable Generation

Let

$$Z_{2jt} = [Z_{2jt,1} \quad Z_{2jt,2} \quad \cdots \quad Z_{2jt,n_{ex}}]'$$

$$\tilde{V}_{jt} = [V_{jt,1} \quad V_{jt,2} \quad \cdots \quad V_{jt,n_{end}} \quad \varepsilon_j]'$$

Then generate  $\{Z_{1t}, Z_{2t}, W_t\}_{t=1}^{ntp}$  from the following distributions

$$[W_{1t} \quad W_{2t} \quad \cdots \quad W_{t,n_{inst}}] \sim N(\mathbf{0}_{n_{inst} \times 1}, CV_{inst})$$

$$[Z'_{21t} \quad Z'_{22t} \quad \cdots \quad Z'_{2n_{cs}t}]' \sim N(\mathbf{0}_{n_{cs} \cdot n_{exo} \times 1}, CV_{ex})$$

$$[\tilde{V}'_{1t} \quad \tilde{V}'_{2t} \quad \cdots \quad \tilde{V}'_{n_{cs}t}]' \sim N(\mathbf{0}_{n_{cs} \cdot (n_{end}+1) \times 1}, CV_{er})$$

## Endogenous Variable and Regressand Generation

- ▶ Randomly Draw  $\alpha_{1d} \in [1, -1]^{n_{exo}}$  for each  $d \in \{1, 2, \dots, n_{end}\}$
- ▶ Randomly Draw  $\alpha_{2d} \in [1, -1]^{n_{tinst}}$  for each  $d \in \{1, 2, \dots, n_{end}\}$
- ▶ Randomly draw a set of integers from  $\mathcal{C}_{n_{cinst}}^{n_{tinst}}$  ways that that you can choose  $n_{cinst}$  instruments from  $n_{tinst}$  total instrument, for each  $j \in \{1, 2, \dots, n_{cs}\}$
- ▶ Map that set of integers to a binary vector  $m_j$  indicating the integers drawn above.
- ▶ Let  $M_j = \text{diag}(m_j)$ , and generate the following

$$Z_{1jd} = \alpha_{0jd} + Z'_{2jt}\alpha_{1d} + W'_t M_j \alpha_{2d} + V_{jt,d} \quad \text{where} \quad \alpha_{0jd} = 1/2 + j/2$$

- ▶ Draw the coefficient vector  $[\beta'_1 \ \beta'_2]' \in [1, -1]^{n_{end} + n_{end}}$ , and generate the following

$$Y_{jt} = \beta_0 + Z'_{1jt}\beta_1 + Z'_{2jt}\beta_2 + e_j + \varepsilon_{jt} \quad \text{where} \quad e_j = 1 + j/2$$

...

