

# Selection of Heterogeneous Instruments in Fixed Effects Panel Regression

Eric Penner, *University of California Berkeley*

August, 2018

In general, the study of panel data is concerned with accommodating the varying sources of heterogeneity that exist between cross sectional units. This heterogeneity may arise when cross sectional units are separated by some spatial distance be it physical distance, difference in cultural norms, or degree of separation in a network of trading partners. Regardless of the source of this heterogeneity the goal is to identify and estimate some set of effects which are common to all cross sections. Most panel data estimators are solely concerned with treating sources of heterogeneity that cannot be excluded from the primary regression in the guise of time or fixed effects. Few if any panel data estimators consider heterogeneity in the relevant instruments needed to treat the presence of endogenous regressors. This paper develops an estimator which accommodates the reasonable assumption that the instruments needed to decompose the variation in one or more endogenous variables will vary by cross section. This paper also incorporates the complicating factor that the instruments relevant to each cross section are an unknown subset of a potentially high dimensional ( $k \gg n$ ) vector.

Motivation for an estimator like the one developed in this paper is most clear in the context of Macroeconomics. There, it is common to run cross industry, or cross country regressions in order to identify some common characteristic like total factor productivity Basu et al. (2006) or growth arising from foreign aid Burnside and Dollar (2000). As is common in economics, in each of the supplied examples the authors acknowledge and treat the presence of endogenous regressors but, both are either unable or unwilling to consider heterogeneous instruments for those endogenous variables. It is well understood that the treatment of endogeneity relies on the use of instrumental variables to decompose the variation of each endogenous regressor into orthogonal parts, one part that is related to the error term and one that isn't. The validity of any treatment of endogeneity relies on the validity of this decomposition. It is reasonable to assume that this decomposition varies by cross section. For example, in the growth and foreign aid panel regression of Burnside and Dollar (2000) foreign aid is the endogenous regressor and it makes sense that the factors which determine the foreign aid that a country like Egypt receives, being that it is neighbors with Israel, is on the Mediterranean, and sits astride the Suez canal, would be different than a country like Papua New Guinea whose foreign aid is likely more humanitarian in nature.

This paper further assumes that the vector of instrumental variables relevant to each cross section is an unknown subset of a large collection. This assumption is really just an acknowledgement that gathering

sufficient domain information regarding each cross section in order to determine what the instruments should be is likely too difficult or not possible. Another aspect of this paper is that this collection of potential instruments can be very high dimensional. This allows the use of more modern and high dimensional data sources, as is encouraged in Varian (2014), to be used for decomposing endogenous regressors, for example; cell phone, and scanner data.

## Methodology

Given a linear in parameters fixed effects panel data model, suppose that for all cross-sections and time periods a known subset of regressors is correlated with the error term. This endogeneity problem will be resolved using the control function approach of Newey et al. (1999), with the additional complication of having the set of instrumental variables which are appropriate/available for at least one cross-section not identical to those available/appropriate for a non empty subset of other cross-sections.

Suppose that for each time period  $t \in \{1, 2, \dots, T\}$ , component of  $Z_{1jt}$   $d \in \{1, 2, \dots, p_1\}$ , and cross-section  $j \in \{1, 2, \dots, q\}$  where  $\{q, T\} \in \mathbb{N}$ ,

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

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

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

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

Where for  $\{p_1, p_2, c, w, w_1, w_2, \dots, w_q\} \subset \mathbb{N}$ ,  $Y_{jt}$ , and  $\varepsilon_{jt}$  are scalar random variables,  $Z_{2jt}$  is a  $p_2$  dimension vector of exogenous random variables,  $Z_{1jt}$  is a vector of endogenous random variables having dimension  $p_1$ .  $e_j$  is a scalar fixed effect,  $\varepsilon_{jt}$  is a scalar error term,  $V_{jdt}$  is a scalar error term.  $W_t = [W_{1t} \ W_{2t} \ \dots \ W_{wt}]'$  is vector of instrumental variables of dimension  $w$ ,  $W_{jt} = [W_{j1t} \ W_{j2t} \ \dots \ W_{jw_{jt}}]'$  is a vector of instrumental variables of dimension  $w_j$  where  $\{W_{jit}\}_{i=1}^{w_j} \subset \{W_{it}\}_{i=1}^w$  and there exists at least one pair  $j, j' \in \{1, 2, \dots, p\}$ , where  $j \neq j'$ , such that  $\{W_{jit}\}_{i=1}^{w_j} \neq \{W_{j'it}\}_{i=1}^{w_{j'}}$ .  $\beta_0$  and  $\alpha_{d0}$  are scalars,  $\beta_1$ ,  $\alpha_{1d}$ ,  $\alpha_{2d}$  are  $p_1 + p_2 = p$ ,  $p_2$ , and  $w_j$  dimensional vectors of real numbers respectively. Lastly for notational convenience let define the following equivalences,

$$Z_{jt} = [Z'_{1jt} \ Z'_{2jt}]' \quad \text{and} \quad V_{jt} = [V_{j1t} \ V_{j2t} \ \dots \ V_{jp_1t}]'$$

### Remarks :

a.) We have in equation (1) a linear in parameters function of primary interest having endogenous random

variable  $Z_{1jt}$ . The focus of this proposal is the identification and estimation of finite dimensional parameter  $\beta_1$ .

As mentioned above  $E(\varepsilon_{jt}|Z_{1jt}) \neq 0$  means that the error term  $\varepsilon_{jt}$  is not orthogonal to the space of functions of  $Z_{1jt}$ . Following the control function approach of Newey et al. (1999), I assume that there exists a random variable  $u_{jt}$  and unknown function  $f_j(\cdot)$  such that;

$$\varepsilon_{jt} = f_j(V_{jt}) + u_{jt} \quad \text{where} \quad E(u_{jt}|\{Z_{jk}, W_{jk}, X_{jk}, V_{jk}\}_{k=1}^t) = 0$$

Furthermore I assume that all unknown functions heretofore define are additive, that is,  $f_j(V_{jt}) = \sum_{d=1}^{p_1} f_{jd}(V_{jdt})$  Where each  $f_{jd}(\cdot)$  is a unknown function. Now where  $t > 1$ , we can rewrite (1) as;

$$Y_{jt} - Y_{j(t-1)} = [Z'_{jt} - Z'_{j(t-1)}]\beta_1 + \sum_{d=1}^{p_1} [f_{jd}(V_{jdt}) - f_{jd}(V_{jd(t-1)})] + u_{jt} - u_{j(t-1)}$$

Adopting the notation that for any finite dimensional random vector  $A_{jt}$  define  $\Delta A_{jt} \equiv A_{jt} - A_{j(t-1)}$  we can further rewrite (1) as;

$$\Delta Y_{jt} = \Delta Z'_{jt}\beta_1 + \sum_{l=1}^{p_1} \Delta f_{jl}(V_{jlt}) + \Delta u_{jt}$$

As a result of the foregoing assumptions and definitions, we can rewrite the model as the following,

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

$$Z_{1jdt} = \alpha_{d0} + Z'_{2jdt}\alpha_{1jd} + W'_{jdt}\alpha_{2jd} + V_{jdt} \tag{2d}$$

$$E(\Delta u_{jt}|\{Z_{jk}, W_{jk}, X_{jk}, V_{jk}\}_{k=1}^t) = 0 \tag{6}$$

$$E(V_{jt}|Z_{2jt}, W_{jt}) = 0 \tag{7}$$

#### Remarks :

- a.) In equation (5), by differencing equation (1), including the control function, and imposing additivity, we have eliminated fixed effect  $e_j$ , but in return we now have a collection of equations  $\{f_{j1}(\cdot), f_{j2}(\cdot), \dots, f_{jp_1}(\cdot)\}$  unique to each cross-section  $j$ , a fate seemingly even worse than the presence of fixed effects.
- b.) In the following I will show that using a variation on the identification technique of Manzan and Zerom (2005), it is possible under mild conditions to identify and estimate finite dimensional parameter  $\beta_1$  and hopefully achieve  $\sqrt{n}$  asymptotic normality without having to jointly estimate  $f_j(V_{jt})$ ,

Next in preparation for the following lemma regarding the identification of parameter vector  $\beta_1$  I define the following density ratios,

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

where for any random vector  $A_j$ ,  $p(A_j)$  is the joint density of that vector. Define following conditional expectations,

$$\begin{aligned} H_{jd}(\Delta Z_{jt}) &= E[\phi_{jt} \Delta Z_{jt} | V_{jdt}, V_{jd(t-1)}] & H_{jd}(\Delta Y_{jt}) &= E[\phi_{jt} \Delta Y_{jt} | V_{jdt}, V_{jd(t-1)}] \\ H_j(\Delta Z_{jt}) &= \sum_{d=1}^{p_1} H_{jd}(\Delta Z_{jt}) & H_j(\Delta Y_{jt}) &= \sum_{d=1}^{p_1} H_{jd}(\Delta Y_{jt}) \end{aligned}$$

Lastly, define the following collection of vectors,

$$\begin{aligned} H(\Delta Y_t) &= [H_1(\Delta Y_{1t}) \ H_2(\Delta Y_{2t}) \ \cdots \ H_q(\Delta Y_{qt})]' & H(\Delta Z_t) &= [H_1(\Delta Z_{1t}) \ H_2(\Delta Z_{2t}) \ \cdots \ H_q(\Delta Z_{qt})]' \\ \Delta Y_t &= [\Delta Y_{1t} \ \Delta Y_{2t} \ \cdots \ \Delta Y_{qt}]' & \Delta Z_t &= [\Delta Z_{1t} \ \Delta Z_{2t} \ \cdots \ \Delta Z_{qt}]' & \Delta u_t &= [\Delta u_{1t} \ \Delta u_{2t} \ \cdots \ \Delta u_{qt}]' \end{aligned}$$

Now that all necessary definitions have been made, the following lemma gives the sufficient conditions for the identification of parameter  $\beta_1$ ,

**Lemma 1.** *Letting  $\phi_t = \text{diag}(\{\phi_{jt}\}_{j=1}^q)$ , if*

- i.)  $E[f_{jd}(V_{jdt})] = E[f_{jd}(V_{jd(t-1)})]$  for all  $d \in \{1, 2, \dots, p_1\}$ ,  $j \in \{1, 2, \dots, q\}$ , and  $t \in \{2, \dots, T\}$
- ii.)  $E\left([\Delta Z_t - H(\Delta Z_t)]' [\Delta Z_t - H(\Delta Z_t)]\right)$  is positive semi definite,

Then  $\beta_1$  is identified, in particular,

$$\beta_1 = E\left([\Delta Z_t - H(\Delta Z_t)]' \phi_t [\Delta Z_t - H(\Delta Z_t)]\right)^{-1} E\left([\Delta Z_t - H(\Delta Z_t)]' \phi_t [\Delta Y_t - H(\Delta Y_t)]\right)$$

*Proof.* I begin with a series of preliminary results, which combine to show the main result. First for  $d \neq k$  consider,

$$\begin{aligned} E[\phi_{jt} \Delta f_{jk}(V_{jkt}) | V_{jdt}, V_{jd(t-1)}] &= \int \phi_{jt} \Delta f_{jk}(V_{jkt}) p(V_{jdt}, V_{jd(t-1)}) p(V_{jdt}, V_{jd(t-1)})^{-1} dV_{j-dt} V_{j-d(t-1)} \\ &= \int \Delta f_{jk}(V_{jkt}) p(V_{jkt}, V_{jk(t-1)}) dV_{jkt} V_{jk(t-1)} \prod_{a \notin \{d, k\}} \int p(V_{jat}, V_{ja(t-1)}) dV_{jat} V_{ja(t-1)} \\ &= E[\Delta f_{jk}(V_{jkt})] = E[f_{jk}(V_{jkt})] - E[f_{jk}(V_{jk(t-1)})] = 0 \end{aligned}$$

Next consider

$$\begin{aligned} E[\phi_{jt}|V_{jdt}, V_{jd(t-1)}] &= \int \phi_{jt} p(V_{jt}, V_{j(t-1)}) p(V_{jdt}, V_{jd(t-1)})^{-1} dV_{j-dt} V_{j-d(t-1)} \\ &= \prod_{a \neq d} \int p(V_{jat}, V_{ja(t-1)}) dV_{jat} V_{ja(t-1)} = 1 \end{aligned}$$

Consequently,

$$\begin{aligned} H_{jd}(\Delta Y_{jt}) &= E[\phi_{jt} \Delta Y_{jt} | V_{jdt}, V_{jd(t-1)}] \\ &= E[\phi_{jt} \Delta Z'_{jt} | V_{jdt}, V_{jd(t-1)}] \beta_1 + \Delta f_{jd}(V_{jdt}) E[\phi_{jt} | V_{jdt}, V_{jd(t-1)}] \\ &\quad + \sum_{k \neq d}^{p_1} E[\phi_{jt} \Delta f_{jk}(V_{jkt}) | V_{jdt}, V_{jd(t-1)}] + E[\phi_{jt} \Delta u_{jt} | V_{jdt}, V_{jd(t-1)}] \\ &= H_{jd}(\Delta Z'_{jt}) \beta_1 + \Delta f_{jd}(V_{jdt}) + E[\phi_{jt} E(\Delta u_{jt} | \{Z_{jk}, W_{jk}, V_{jk}\}_{k=1}^t) | V_{jdt}, V_{jd(t-1)}] \\ &= H_{jd}(\Delta Z'_{jt}) \beta_1 + \Delta f_{jd}(V_{jdt}) \end{aligned}$$

Furthermore

$$H_j(\Delta Y_{jt}) = \sum_{d=1}^{p_1} H_{jd}(\Delta Y_{jt}) = \sum_{d=1}^{p_1} H_{jd}(\Delta Z'_{jt}) \beta_1 + \sum_{d=1}^{p_1} \Delta f_{jd}(V_{jdt}) = H_j(\Delta Z_{jt}) \beta_1 + \sum_{d=1}^{p_1} \Delta f_{jd}(V_{jdt})$$

Consequently

$$\sqrt{\phi_{jt}}[\Delta Y_{jt} - H_j(\Delta Y_{jt})] = \sqrt{\phi_{jt}}[\Delta Z_{jt} - H_j(\Delta Z_{jt})] \beta_1 + \sqrt{\phi_{jt}} \Delta u_{jt}$$

and stacking these equations into a vector,

$$\sqrt{\phi_t}[\Delta Y_t - H(\Delta Y_t)] = \sqrt{\phi_t}[\Delta Z_t - H(\Delta Z_t)] \beta_1 + \sqrt{\phi_t} \Delta u_t$$

The result follows since  $E[\sqrt{\phi_t} E(\Delta u_t | \{Z_{jk}, W_{jk}, V_{jk}\}_{k=1}^t)] = 0$  □

## Estimation

### Secondary Equations

The specification of equation (2d) is quite general where; intercepts  $\alpha_{0jd}$ , and coefficients on exogenous regressors  $\alpha_{1jd}$  are unique to each cross-section. Furthermore, although instruments  $W_{jt}$  are shared across cross-sections their coefficients  $\alpha_{2jd}$  are unique, and as yet there is no sense in which error terms

$V_{jdt}$  are correlated across cross-section. As a result there are a number of restrictions on the regressors and parameters of equation (2d) which can be imposed and have a substantial effect on the manner in which  $V_{jdt}$  will be estimated.

**Note:** For all vectors of dimension greater than two, I reference the individual elements of each vector by including a comma followed by a scalar value in the subscript. I reference the entire vector whenever the comma is omitted. For example,

$$W_{jt} = \begin{bmatrix} W_{jt,1} & W_{jt,2} & \cdots & W_{jt,w_j} \end{bmatrix}'$$

**Case 1:**  $W_{jt}$  is an known subset of  $W_t$ .

If so, estimation is comprised of  $q$  separate OLS regressions.

$$(\hat{\alpha}_{0jd}, \hat{\alpha}_{1jd}, \hat{\alpha}_{2jd}) = \arg \min \sum_{t=1}^T (Z_{1jdt} - \alpha_0 - Z'_{2jt}\alpha_1 - W'_{jt}\alpha_2)^2$$

where  $\alpha_0 \in \mathbb{R}$ ,  $\alpha_1 \in \mathbb{R}^{p_2}$ , and  $\alpha_2 \in \mathbb{R}^{w_j}$ . So that

$$\hat{V}_{jdt} = Z_{1jdt} - \hat{\alpha}_{0jd} - Z'_{2jt}\hat{\alpha}_{1jd} - W'_{jt}\hat{\alpha}_{2jd}$$

**Case 2:**  $W_{jt}$  is an unknown subset of  $W_t$ .

If so, estimation is comprised of  $q$  separate regressions, each of which will incorporate a subset selection routine. Let  $\alpha_{2jd} = [\alpha_{2jd,1} \ \alpha_{2jd,2} \ \cdots \ \alpha_{2jd,w}]$  where  $\alpha_{2jd,l} = 0$  whenever  $W_{t,l} \notin W_{jt}$ . To facilitate subset selection I apply the lasso estimator by imposing an  $\ell^1$  penalty on estimated coefficients  $\hat{\alpha}_{2jd}$ .

$$(\hat{\alpha}_{0jd}, \hat{\alpha}_{1jd}, \hat{\alpha}_{2jd}) = \arg \min \sum_{t=1}^T (Z_{1jdt} - \alpha_0 - Z'_{2jt}\alpha_1 - W'_t\alpha_2)^2 \quad \text{subject to} \quad \sum_{l=1}^w |a_{2,l}| \leq \lambda$$

where  $\alpha_0 \in \mathbb{R}$ ,  $\alpha_1 \in \mathbb{R}^{p_2}$ , and  $\alpha_2 \in \mathbb{R}^w$ . So that again,

$$\hat{V}_{jdt} = Z_{1jdt} - \hat{\alpha}_{0jd} - Z'_{2jt}\hat{\alpha}_{1jd} - W'_t\hat{\alpha}_{2jd}$$

**Case 3:**  $W_{jt}$  is an known subset of  $W_t$ ,  $\alpha_{1jd} = \alpha_{1j'd} \equiv \alpha_{1d}$  for all  $j, j' \in \{1, 2, \dots, q\}$ , and  $\alpha_{2jd,l} = \alpha_{2j'd,l} \equiv \alpha_{2d,l}$  whenever  $W_{t,l} \in W_{jt}$  and  $W_{t,l} \in W_{j't}$ . Let

$$1[W_{t,l} \in W_{tj}] = \begin{cases} 1 & \text{if } W_{t,l} \in W_{tj} \\ 0 & \text{otherwise} \end{cases}$$

and  $M_j = \text{diag}(\{1[W_{tl} \in W_{jt}]\}_{l=1}^w)$  so that,

$$\begin{aligned} Z_{1jdt} &= \alpha_{0jd} + Z'_{2jt}\alpha_{1jd} + W'_{jt}\alpha_{2jd} + V_{jdt} \\ &= \alpha_{0jd} + Z'_{2jt}\alpha_{1d} + W'_t M_j \alpha_{2d} + V_{jdt} \end{aligned}$$

Now let  $\Delta Z_{1jdt} = Z_{1jdt} - Z_{1jd(t-1)}$ ,  $\Delta Z_{2jdt} = Z_{2jdt} - Z_{2jd(t-1)}$ ,  $\Delta W_t = W_t - W_{t-1}$ , and  $\Delta V_{jdt} = V_{jdt} - V_{jd(t-1)}$  so that,

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

As a result,

$$(\hat{\alpha}_{1d}, \hat{\alpha}_{2d}) = \arg \min \sum_{j=1}^q \sum_{t=1}^T (\Delta Z_{ijdt} - \Delta Z'_{2jt}\alpha_1 - \Delta W'_t M_j \alpha_2)^2$$

so that given

$$\alpha_{0jd} = E(V_{jdt} + \alpha_{0jd}) = E(Z_{1jdt} - Z'_{2jt}\alpha_{1d} - W'_t M_j \alpha_{2d})$$

we have

$$\hat{V}_{jdt} = Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{1d} - W'_t M_j \hat{\alpha}_{2d} - T^{-1} \sum_{t=1}^T (Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{1d} - W'_t M_j \hat{\alpha}_{2d})$$

.

**Case 4:**  $W_{jt}$  is an unknown subset of  $W_t$ ,  $\alpha_{1jd} = \alpha_{1j'd} \equiv \alpha_{1d}$  for all  $j, j' \in \{1, 2, \dots, q\}$ , and  $\alpha_{2jd,l} = \alpha_{2j'd,l} \equiv \alpha_{2d,l}$  whenever  $W_{t,l} \in W_{jt}$  and  $W_{t,l} \in W_{j't}$ . Inheriting notation from case 3 we again have,

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

In order to introduce our selection procedure we will estimate the coefficients on  $W_t$  as if that are not identical, then average the non zero estimates to construct a single estimate. Consider,

$$(\hat{\alpha}_{1d}, \hat{\alpha}_{2d}) = \arg \min \sum_{j=1}^q \sum_{t=2}^T (\Delta Z_{1jdt} - \Delta Z'_{2jt}\alpha_1 - \Delta W'_t \alpha_{2j})^2 \quad \text{subject to} \quad \sum_{l=1}^w |\alpha_{2j,l}| \leq \lambda \quad \text{for all } 1 \leq j \leq q$$

Consequently define for some  $\varepsilon > 0$

$$\tilde{\alpha}_{2d,l} = \frac{\sum_{l=1}^q \hat{\alpha}_{2jd,l} 1[\hat{\alpha}_{2jd,l} > \varepsilon]}{\sum_{l=1}^q 1[\hat{\alpha}_{2jd,l} > \varepsilon]}$$

Now let  $\tilde{\alpha}_{2d} = [\tilde{\alpha}_{2d,1} \ \tilde{\alpha}_{2d,2} \ \cdots \ \tilde{\alpha}_{2d,w}]'$  and  $\tilde{M}_{jd} = \text{diag}(\{1[\hat{\alpha}_{2jd,l} > \varepsilon]\}_{l=1}^w)$  so that,

$$\hat{V}_{jdt} = Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{1d} - W'_t\tilde{M}_{jd}\tilde{\alpha}_{2d} - T^{-1} \sum_{t=1}^T (Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{1d} - W'_t\tilde{M}_{jd}\tilde{\alpha}_{2d})$$

## Primary Equation

Now having estimates of all error terms, I construct the following estimates of the primary equation coefficients in the following way.

### Density Estimation:

Obtain Rosenblatt Kernel Density Estimates of,  $p(V_{jdt}, V_{jdc})$ , and  $p(V_{jdt}, V_{jct})$  using  $\{\hat{V}_{jdt}, \hat{V}_{jct}\}_{t=2}^T$ .

$$\begin{aligned} \hat{p}(\hat{V}_{jdt,l}, \hat{V}_{jd(t-1),l}) &= (nh_1h_2)^{-1} \sum_{i=2}^T k_1[h_1^{-1}(\hat{V}_{jdi,l} - \hat{V}_{jdt,l})]k_2[h_2^{-1}(\hat{V}_{jdi,l} - \hat{V}_{jd(t-1),l})] \\ \hat{p}(\hat{V}_{jdt,l}, \hat{V}_{j(t-1),l}) &= (nh_1^{p_1}h_2^{p_1})^{-1} \sum_{i=2}^T \prod_{d=1}^{p_1} k_1[h_1^{-1}(\hat{V}_{jdi,l} - \hat{V}_{jdt,l})]k_2[h_2^{-1}(\hat{V}_{jdi,l} - \hat{V}_{jd(t-1),l})] \end{aligned}$$

Form estimated density ratio  $\hat{\phi}_{jt}$  with densities estimated in previous step.

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

### H Function Estimation:

Now estimate the conditional expectation functions by a variation on a Nadaraya Watson estimation.

$$\hat{H}_{j,d}(\Delta Z_{jt,a}) = [(T-1)b_1b_2]^{-1} \sum_{l \neq t, l' > 1}^T k_1[b_1^{-1}(V_{jlt,d} - V_{jdt,d})]k_2[b_2^{-1}(V_{j(l-1),d} - V_{j(t-1),d})]\hat{\theta}_{jl,d}\Delta Z_{jl,a}$$

then construct

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

where  $\Delta Z_{jt} = [\Delta Z_{jt,1} \ \Delta Z_{jt,2} \ \cdots \ \Delta Z_{jt,(p_1+p_2)}]'$ . Then construct the following vectors.

$$\begin{aligned} \hat{H}(\Delta Y_t) &= [\hat{H}_1(\Delta Y_{1t}) \ \hat{H}_2(\Delta Y_{2t}) \ \cdots \ \hat{H}_q(\Delta Y_{qt})]' \\ \hat{H}(\Delta Z_t) &= [\hat{H}_1(\Delta Z_{1t}) \ \hat{H}_2(\Delta Z_{2t}) \ \cdots \ \hat{H}_q(\Delta Z_{qt})]' \\ \Delta Z &= [\Delta Z'_2 \ \Delta Z'_3 \ \cdots \ \Delta Z'_T]' \\ \Delta Y &= [\Delta Y_2 \ \Delta Y_3 \ \cdots \ \Delta Y_T]' \\ \hat{H}(\Delta Z) &= [\hat{H}(\Delta Z_2)' \ \hat{H}(\Delta Z_3)' \ \cdots \ \hat{H}(\Delta Z_T)'] \end{aligned}$$



$$\hat{H}(\Delta Y) = [ \hat{H}(\Delta Y_2) \ \hat{H}(\Delta Y_3) \ \cdots \ \hat{H}(\Delta Y_T) ]$$

$\beta_1$  **Coefficient Estimation:**

Let  $\hat{\phi} = \text{diag}(\{\hat{\phi}_t\}_{c+1}^T)$  and calculate  $\hat{\beta}_1$  as follows,

$$\hat{\beta}_1 = \left( [\Delta Z - \hat{H}(\Delta Z)]' \hat{\phi} [\Delta Z - \hat{H}(\Delta Z)] \right)^{-1} \left( [\Delta Z - \hat{H}(\Delta Z)]' \hat{\phi} [\Delta Y - \hat{H}(\Delta Y)] \right)$$

## Monte Carlo Exercise

Please see accompanying file

`shi_monte_carlo.html`

## References

- Basu, S., Fernald, J. G., Kimball, M. S., December 2006. Are technology improvements contractionary. *American Economic Review* 96 (5), 1418–1448.
- Burnside, C., Dollar, D., 868 2000. Aid , policies, and growth. *American Economic Review* 90 (4), 847.
- Manzan, S., Zerom, D., 2005. Kernel estimation of a partially linear model. *Statistics and Probability Letters* 73, 313 – 322.
- Newey, W. K., Powell, J. L., Vella, F., May 1999. Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67 (3), 565–603.
- Varian, H., 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28 (2), 3–28.