Proposal: Parametric Panel, Selection, and Control

by Eric Penner

Introduction

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, Powell, and Vella (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. That is, suppose that for each time period $t \in \{1, 2, \ldots, T\}$, component of Z_{1jt} $d \in \{1, 2, \ldots, p_1\}$, and cross-section $j \in \{1, 2, \ldots, q\}$ where $\{q, T\} \in \mathbb{N}$,

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

$$Z_{1jdt} = \alpha_{jd0} + [\ Z_{1jdt-1} \ \ldots \ Z_{1jdt-c}\ ]\alpha_{jd1} + Z'_{2jt}\alpha_{jd2} + W'_{jt}\alpha_{jd3} + V_{jdt} \tag{2d}$$

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

$$E(V_{jdt}|Z_{1jt-1}, \ldots, Z_{1jt-c_1}, Z_{2jt}, W_{jt}) = 0 \tag{4}$$

Where for $\{p_1, p_2, c, w, w_1, w_2, \ldots, w_q\} \subset \mathbb{N}$, Y_{jt}, and ε_{jt} are scalar random variables, Z_{2jt} is a p_2 dimesion 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, ε_{jt} is a scalar error term, V_{jdt} is a scalar error term. $W_t = [\ W_{1t} \ \ W_{2t} \ \ \cdots \ \ W_{wt}\ ]'$ is vector of instrumental variables of dimension w, $W_{jt} = [\ W_{j1t} \ \ W_{j2t} \ \ \cdots \ \ W_{jw_jt}\ ]'$ 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 $j, j' \in \{1, 2, \ldots, p\}$, where $j \neq j'$, such that $\{W_{jit}\}_{i=1}^{w_j} \neq \{W_{j'it}\}_{i=1}^{w_{j'}}$. β_0 and α_{d0} are scalars, β_1, α_{d1}, α_{d2}, α_{d3} are $p_1 + p_2 = p$, c, p_2, and w_j dimensional vectors of real numbers respectively. Lastly for notational convenience let define the following equivalences,

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

Remarks :

a.) We have in equation (1) a linear in parameters function of primary interest having endogenous random variable Z_{2jt}. The focus of this proposal is the identification and estimation of finite dimensional parameter β_1.

b.) In regards to the underlying stochastic process at work, the current level of generalization allows for either the assumption that, $\alpha_{d1}=0$, and $\{Z_{2jt}, W_t, V_{jt}, \varepsilon_{jt}\}_{t=1}^{T}$ is an i.i.d sequence of random vectors. Or that $\{Z_{2jt}, W_t, V_{jt}\}_{t=1}^{T}$ is a stationary mixing sequence of d.i.d random vectors and $\{\varepsilon_{jt}\}_{t=1}^{T}$ is i.i.d or i.n.i.d.

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

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

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 $t > c$, we can rewrite (1) as;

$$Y_{jt} - Y_{jc} = [Z'_{jt} - Z'_{jc}]\beta_1 + \sum_{d=1}^{p_1}[f_{jd}(V_{jdt}) - f_{jd}(V_{jdc})] + u_{jt} - u_{jc}$$

Adopting the notation that for any finite dimensional random vector A_{jt} define $\Delta A_{jt} \equiv A_{jt} - A_{jc}$ 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,

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

$$Z_{1jdt} = \alpha_{d0} + [\ Z'_{1jdt-1}\ \dots\ Z'_{1jdt-c}\ ]\alpha_{d1} + Z'_{2jt}\alpha_{d2} + W'_{jt}\alpha_{d3} + V_{jdt} \tag{2d}$$

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

$$E(V_{jt}|Z_{1jt-1}, \dots, Z_{1jt-c_1}, Z_{2jt}, W_{jt}) = 0 \tag{7}$$

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) SaPL, it is possible under mild conditions to identify and estimate finite dimensional parameter β_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 β_1 I define the following density ratios,

$$\phi_{jt} = \frac{\prod_{d=1}^{p_1} p(V_{jdt}, V_{jdc})}{p(V_{jt}, V_{jc})}$$

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

$$H_{jd}(\Delta Z_{jt}) = E[\phi_{jt}\Delta Z_{jt}|V_{jdt}, V_{jdc}] \qquad H_{jd}(\Delta Y_{jt}) = E[\phi_{jt}\Delta Y_{jt}|V_{jdt}, V_{jdc}]$$
$$H_j(\Delta Z_{jt}) = \sum_{d=1}^{p_1} H_{jd}(\Delta Z_{jt}) \qquad H_j(\Delta Y_{jt}) = \sum_{d=1}^{p_1} H_{jd}(\Delta Y_{jt})$$

Lastly, define the following collection of vectors,

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

Now that all necessary definitions have been made, the following lemma gives the sufficient conditions for the identification of parameter β_1,

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

i.) $E\big[f_{jd}(V_{jdt})\big] = E\big[f_{jd}(V_{jdc})\big]$ *for all* $d \in \{1, 2, \ldots, p_1\}$, $j \in \{1, 2, \cdots, q\}$, *and* $t \in \{c+1, \ldots, T\}$

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

Then β_1 *is identified, in particular,*

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

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

$$E\big[\phi_{jt}\Delta f_{jk}(V_{jkt})|V_{jdt}, V_{jdc}\big] = \int \phi_{jt}\Delta f_{jk}(V_{jkt})p(V_{jt}, V_{jc})p(V_{jdt}, V_{jdc})^{-1}dV_{j-dt}V_{j-dc}$$
$$= \int \Delta f_{jk}(V_{jkt})p(V_{jkt}, V_{jkc})dV_{jkt}V_{jkc} \prod_{a\notin\{d,k\}} \int p(V_{jat}, V_{jac})dV_{jat}V_{jac}$$

$$= E\big[\Delta f_{jk}(V_{jkt})\big] = E\big[f_{jk}(V_{jkt})\big] - E\big[f_{jk}(V_{jkc})\big] = 0$$

Next consider

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

Consequently,

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

Furthermore

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

Consequently

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

and stacking these equations into a vector,

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

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

Estimation

Step One:

- Case 1: (Selection), For all $j \in \{1, 2, \ldots, q\}$, W_{jt} is an unknown subset of W_t, consequently equations (2d) is estimated with the incorporation of a variable selection procedure as a first/preliminary step.

- Case 2: (No selection), For all $j \in \{1, 2, \ldots, q\}$, W_{jt} is a known subset of W_t, consequently equations (2d) are estimated parametrically.

- In either case the estimation of (2d) will generate residual vectors;

$$\hat{V}_t = [\ \hat{V}_{1t}' \ \ \hat{V}_{2t}' \ \ \cdots \ \ \hat{V}_{p_1 t}'\ ]'$$

Step Two:

- 2.1 : Obtain Rosenblatt Kernel Density Estimates of, $p(V_{jdt}, V_{jdc})$, and $p(V_{jt}, V_{jc})$ using $\{\hat{V}_{jt}, \hat{V}_{jc}\}_{t=c+1}^{T}$.

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

Step Three:

- 3.1: Obtain Nadaraya Watson estimates of $H_{jd}(\Delta Z_{jt})$, and $H_{jd}(\Delta Y_{jt})$, then construct

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

- 3.2: Let $T^* = T - c - 1$ and construct the following vectors using estimates from previous steps.

$$\begin{aligned}
\hat{H}(\Delta Y_t) &= [\ \hat{H}_1(\Delta Y_{1t}) \ \ \hat{H}_2(\Delta Y_{2t}) \ \ \cdots \ \ \hat{H}_q(\Delta Y_{qt})\ ]' \\
\hat{H}(\Delta Z_t) &= [\ \hat{H}_1(\Delta Z_{1t}) \ \ \hat{H}_2(\Delta Z_{2t}) \ \ \cdots \ \ \hat{H}_q(\Delta Z_{qt})\ ]' \\
\Delta Z &= [\ \Delta Z_{c+1}' \ \ \Delta Z_{c+2}' \ \ \cdots \ \ \Delta Z_T'\ ]' \\
\Delta Y &= [\ \Delta Y_{c+1} \ \ \Delta Y_{c+2} \ \ \cdots \ \ \Delta Y_T\ ]' \\
\hat{H}(\Delta Z) &= \Big[\ \hat{H}(\Delta Z_{c+1})' \ \ \hat{H}(\Delta Z_{c+2})' \ \ \cdots \ \ \hat{H}(\Delta Z_T)'\ \Big] \\
\hat{H}(\Delta Y) &= \Big[\ \hat{H}(\Delta Y_{c+1}) \ \ \hat{H}(\Delta Y_{c+2}) \ \ \cdots \ \ \hat{H}(\Delta Y_T)\ \Big]
\end{aligned}$$

Step Four:

- 4.1: Let $\hat{\phi} = diag\big(\{\hat{\phi}_t\}_{c+1}^{T}\big)$ Calculate $\hat{\beta}_1$ as follows,

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

Estimation of Secondary Equations

The specification of eqn (2d) is quite general; intercepts α_{jd0}, coefficients on lagged endogenous regressors α_{jd1}, and coefficients on exogenous regressors α_{jd2} are unique to each cross-section. Furthermore, although instruments W_{jt} are shared across cross-sections their coefficients α_{jd3} are unique, and 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 eqn (2d) which have a substantial effect on the manner in which they will be estimated.

Note: In order to drastically simplify estimation, for the time being, I assume that $\alpha_{jd} = \vec{0}$ eliminating the need to deal with any time series concerns.

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

If so estimation is comprised of q separate OLS regressions.

$$(\hat{\alpha}_{0jd}, \hat{\alpha}_{2jd}, \hat{\alpha}_{3jd}) = \arg\min \sum_{t=1}^{T} \left(Z_{ijt} - \alpha_0 - Z'_{2jt}\alpha_2 - W'_{jt}\alpha_3\right)^2$$

where $\alpha_0 \in \mathbb{R}$, $\alpha_1 \in \mathbb{R}^{p_2}$, and $\alpha_3 \in \mathbb{R}^{w_j}$. So that $\hat{V}_{jdt} = Z_{1jdt} - \hat{\alpha}_{0jd} - Z'_{2jt}\hat{\alpha}_{2jd} - W'_{jt}\hat{\alpha}_{3jd}$.

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. In particular, here I will utilize the lasso estimator. In this case let $\alpha_{3jd} = [\ \alpha_{3jd1} \ \ \alpha_{3jd2} \ \ \cdots \ \ \alpha_{3jdw}\ ]$ where $\alpha_{jdl} = 0$ whenever W_{tl} is not an element of W_{jt}

$$(\hat{\alpha}_{0jd}, \hat{\alpha}_{2jd}, \hat{\alpha}_{3jd}) = \arg\min \sum_{t=1}^{T} \left(Z_{ijt} - \alpha_0 - Z'_{2jt}\alpha_2 - W'_{t}\alpha_3\right)^2 \quad \text{subject to} \quad \sum_{l=1}^{w} |a_{3l}| \leq t$$

where $\alpha_0 \in \mathbb{R}$, $\alpha_1 \in \mathbb{R}^{p_2}$, and $\alpha_3 \in \mathbb{R}^{w}$. So that $\hat{V}_{jdt} = Z_{1jdt} - \hat{\alpha}_{0jd} - Z'_{2jt}\hat{\alpha}_{2jd} - W'_{t}\hat{\alpha}_{3jd}$.

Case 3: W_{jt} is an known subset of W_t, $\alpha_{2jd} = \alpha_{2d}$ for all $j \in \{1, 2, \ldots, q\}$, and $\alpha_{3jdl} = \alpha_{3dl}$ whenever W_{tl} is an element of W_{jt}. Let $M_j = diag(\{1\{W_{tl} \in W_{jt}\}\}_{l=1}^{w})$ so that,

$$\begin{aligned} Z_{1jdt} &= \alpha_{0jd} + Z'_{2jt}\alpha_{jd2} + W'_{jt}\alpha_{jd3} + V_{jdt} \\ &= \alpha_{0jd} + Z'_{2jt}\alpha_{d2} + W'_{t}M_j\alpha_{d3} + V_{jdt} \end{aligned}$$

Now let $\Delta Z_{1jdt} = Z_{1jdt} - Z_{1jdt-1}$, $\Delta Z_{2jdt} = Z_{2jdt} - Z_{2jdt-1}$, $\Delta W_t = W_t - W_{t-1}$, and $\Delta V_{jdt} = V_{jdt} - V_{jdt-1}$

so that,

$$\Delta Z_{1jdt} = \Delta Z'_{2jt}\alpha_{d2} + \Delta W'_t M_j \alpha_{d3} + \Delta V_{jdt}$$

As a result,

$$(\hat{\alpha}_{2d}, \hat{\alpha}_{3d}) = \arg\min \sum_{j=1}^{q}\sum_{t=1}^{T}\left(\Delta Z_{ijt} - \Delta Z'_{2jt}\alpha_2 - \Delta W'_t M_j \alpha_3\right)^2$$

and $\hat{V}_{jdt} = Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{2d} - W'_{jt}\hat{\alpha}_{3d} - T^{-1}\sum_{t=1}^{T}(Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{2d} - W'_t M_j \hat{\alpha}_{3d})$. Since $E(Z_{1jdt} - Z'_{2jt}\alpha_{2d} - W'_{jt}\alpha_{3d}) = E(V_{jdt} + \alpha_{0jd}) = \alpha_{0jd}$.

Case 4: W_{jt} is an unknown subset of W_t, $\alpha_{2jd} = \alpha_{2d}$ for all $j \in \{1, 2, \ldots, q\}$, and $\alpha_{3jdl} = \alpha_{3dl}$ whenever W_{tl} is an element of W_{jt}. Let $M_j = diag(\{1\{W_{tl} \in W_{jt}\}\}_{l=1}^{w})$ so that as before,

$$\Delta Z_{1jdt} = \Delta Z'_{2jt}\alpha_{d2} + \Delta W'_t M_j \alpha_{d3} + \Delta V_{jdt}$$

but 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,

$$(\hat{\alpha}_{2d}, \hat{\alpha}_{3d}) = \arg\min \sum_{j=1}^{q}\sum_{t=1}^{T}\left(\Delta Z_{ijt} - \Delta Z'_{2jt}\alpha_2 - \Delta W'_t \alpha_{3j}\right)^2 \quad \text{subject to} \quad \sum_{l=1}^{w} |\alpha_{3jl}| \leq t \quad \text{for all } 1 \leq j \leq q$$

Where $\hat{\alpha}_{3d} = [\ \hat{\alpha}'_{31d}\ \ \hat{\alpha}'_{32d}\ \ \cdots\ \ \hat{\alpha}'_{3qd}\ ]'$. Consequently define for some $\varepsilon > 0$

$$\tilde{\alpha}_{3dl} = \frac{\sum_{j=1}^{q}\hat{\alpha}_{3jdl}1\{\hat{\alpha}_{3jdl} > \varepsilon\}}{\sum_{j=1}^{q}1\{\hat{\alpha}_{3jdl} > \varepsilon\}}$$

Now let $\tilde{\alpha}_{3d} = [\ \tilde{\alpha}_{3d1}\ \ \tilde{\alpha}_{3d2}\ \ \cdots\ \ \tilde{\alpha}_{3dw}\ ]'$ and $\tilde{M}_{jd} = diag(\{1\{\hat{\alpha}_{3jdl} > \varepsilon\}\}_{l=1}^{w})$ so that,

$$\hat{V}_{jdt} = Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{2d} - W'_t\tilde{M}_{jd}\tilde{\alpha}_{3d} - T^{-1}\sum_{t=1}^{T}(Z_{1jdt} - Z'_{2jt}\hat{\alpha}_{2d} - W'_t\tilde{M}_{jd}\tilde{\alpha}_{3d})$$