SUPPLEMENTARY APPENDIX

Identifying Distributional Characteristics in Random Coefficients Panel Data Models

by Manuel Arellano and Stéphane Bonhomme

This appendix contains: identification and estimation results on higher-order moments (Section A), some results on efficiency bounds calculations (Sections B and C) and the computation of asymptotic standard errors (Section D); additional computations for the examples treated in the text (Section E); an outline of a test of the covariance structure of errors (Section F); the calculation of the order conditions, for restrictions in levels and within restrictions (Section G); and a short review of some properties of characteristic functions and cumulants (Section H).

A Higher-order moments

In applications, it may be of interest to document the skewness and kurtosis of individual effects in addition to mean and variance. The restrictions on log-characteristic functions may be used to derive restrictions on higher-order moments of effects and errors.

A.1 Identification

Some notation will be useful. Let $\mathbf{U}$ be an n-dimensional random vector with zero mean and well-defined moments to the fourth-order. We denote by $\boldsymbol{\kappa}_3(\mathbf{U})$ the n^3-dimensional cumulant vector of order 3 whose elements $\kappa_3^{i,j,k}(\mathbf{U})$, for $(i,j,k) \in \{1,...,n\}^3$, are arranged in lexicographic order. Likewise, we denote by $\boldsymbol{\kappa}_4(\mathbf{U})$ the vector of n^4 cumulants $\kappa_4^{i,j,k,\ell}(\mathbf{U})$ of order 4.

Taking third and fourth derivatives at the origin in (40), we obtain the following restrictions:[1]

$$\begin{aligned} \boldsymbol{\kappa}_3\left(\mathbf{y}_i|\mathbf{W}_i\right) &= \left(\mathbf{X}_i \otimes \mathbf{X}_i \otimes \mathbf{X}_i\right) \boldsymbol{\kappa}_3\left(\boldsymbol{\gamma}_i|\mathbf{W}_i\right) + \boldsymbol{\kappa}_3\left(\mathbf{v}_i|\mathbf{W}_i\right), & \text{(A1)} \\ \boldsymbol{\kappa}_4\left(\mathbf{y}_i|\mathbf{W}_i\right) &= \left(\mathbf{X}_i \otimes \mathbf{X}_i \otimes \mathbf{X}_i \otimes \mathbf{X}_i\right) \boldsymbol{\kappa}_4\left(\boldsymbol{\gamma}_i|\mathbf{W}_i\right) + \boldsymbol{\kappa}_4\left(\mathbf{v}_i|\mathbf{W}_i\right). & \text{(A2)} \end{aligned}$$

The MA restrictions on errors (Assumption 5) implies that error cumulants satisfy restrictions of the form:

$$\begin{aligned} \boldsymbol{\kappa}_3\left(\mathbf{v}_i|\mathbf{W}_i\right) &= \mathbf{S}_3\boldsymbol{\omega}_{3i}, & \text{(A3)} \\ \boldsymbol{\kappa}_4\left(\mathbf{v}_i|\mathbf{W}_i\right) &= \mathbf{S}_4\boldsymbol{\omega}_{4i}, & \text{(A4)} \end{aligned}$$

where $\mathbf{S}_3$ and $\mathbf{S}_4$ are selection matrices and $\boldsymbol{\omega}_{3i}$ and $\boldsymbol{\omega}_{4i}$ are vectors that may depend on $\mathbf{W}_i$. Under these assumptions, identification of cumulants can be shown if rank conditions analogous to (24) are satisfied. In addition, the arguments of Subsection 3.3 can be extended to compute the semiparametric information bound for higher-order moments jointly with means, variances, and common parameters. See Section C below.

[1] The validity of (A1) and (A2) relies on the assumption that third- and fourth-order conditional cumulants are finite. Theorem 2, in contrast, only required the existence of second-order moments.

A.2 Estimation

The above identification analysis directly suggests an estimation approach for conditional higher order cumulants of error terms and fixed effects. Using moment restrictions in levels (A1) and (A2), together with the independent moving-average restrictions (A3) and (A4), we see that the vectors of third- and fourth-order conditional cumulants of errors can be estimated as:

$$\begin{aligned}\widehat{\boldsymbol{\kappa}}_3\left(\mathbf{v}_i|\mathbf{W}_i\right) &= \mathbf{S}_3\left[\mathbf{M}_i^{(3)}\mathbf{S}_3\right]^{\dagger}\mathbf{M}_i^{(3)}\widehat{\boldsymbol{\kappa}}_3\left(\mathbf{y}_i|\mathbf{W}_i\right),\\ \widehat{\boldsymbol{\kappa}}_4\left(\mathbf{v}_i|\mathbf{W}_i\right) &= \mathbf{S}_4\left[\mathbf{M}_i^{(4)}\mathbf{S}_4\right]^{\dagger}\mathbf{M}_i^{(4)}\widehat{\boldsymbol{\kappa}}_4\left(\mathbf{y}_i|\mathbf{W}_i\right),\end{aligned}$$

where $\mathbf{M}_i^{(3)}$ and $\mathbf{M}_i^{(4)}$ are analogs of $\mathbf{M}_i$ for third- and fourth-order restrictions, respectively. For example, $\mathbf{M}_i^{(3)}$ satisfies $\mathbf{M}_i^{(3)}\left(\mathbf{X}_i\otimes\mathbf{X}_i\otimes\mathbf{X}_i\right)=\mathbf{0}$. In addition, $\widehat{\boldsymbol{\kappa}}_3\left(\mathbf{y}_i|\mathbf{W}_i\right)$ and $\widehat{\boldsymbol{\kappa}}_4\left(\mathbf{y}_i|\mathbf{W}_i\right)$ denote nonparametric estimates of the conditional cumulants of the data.

Third- and fourth-order conditional cumulants of individual effects can be estimated by:

$$\begin{aligned}\widehat{\boldsymbol{\kappa}}_3\left(\boldsymbol{\gamma}_i|\mathbf{W}_i,\mathbb{S}\right) &= \widehat{\boldsymbol{\kappa}}_3\left(\widehat{\boldsymbol{\gamma}}_i|\mathbf{W}_i,\mathbb{S}\right)-\left(\mathbf{H}_i\otimes\mathbf{H}_i\otimes\mathbf{H}_i\right)\mathbf{S}_3\left[\mathbf{M}_i^{(3)}\mathbf{S}_3\right]^{\dagger}\mathbf{M}_i^{(3)}\widehat{\boldsymbol{\kappa}}_3\left(\mathbf{y}_i|\mathbf{W}_i,\mathbb{S}\right),\\ \widehat{\boldsymbol{\kappa}}_4\left(\boldsymbol{\gamma}_i|\mathbf{W}_i,\mathbb{S}\right) &= \widehat{\boldsymbol{\kappa}}_4\left(\widehat{\boldsymbol{\gamma}}_i|\mathbf{W}_i,\mathbb{S}\right)-\left(\mathbf{H}_i\otimes\mathbf{H}_i\otimes\mathbf{H}_i\otimes\mathbf{H}_i\right)\mathbf{S}_4\left[\mathbf{M}_i^{(4)}\mathbf{S}_4\right]^{\dagger}\mathbf{M}_i^{(4)}\widehat{\boldsymbol{\kappa}}_4\left(\mathbf{y}_i|\mathbf{W}_i,\mathbb{S}\right),\end{aligned}$$

where $\widehat{\boldsymbol{\kappa}}_3\left(\widehat{\boldsymbol{\gamma}}_i|\mathbf{W}_i,\mathbb{S}\right)$ and $\widehat{\boldsymbol{\kappa}}_4\left(\widehat{\boldsymbol{\gamma}}_i|\mathbf{W}_i,\mathbb{S}\right)$ are nonparametric estimates of the conditional cumulants of the fixed-effects estimates.

Note that estimates of unconditional cumulants can be obtained using the above. However, the resulting estimators will depend on nonparametric estimates of conditional quantities, unlike the situation for unconditional means and variances.

Example 2. Consider again Example 2 with $L=3$, for a sequence of covariates $s_{i1}=1$, $s_{i2}=0$, $s_{i3}=0$. Assume in addition that $v_{i\ell}$ are i.i.d. Using the within information, only the moments of $v_{i3}-v_{i2}=y_{i3}-y_{i2}$ are identified. So the third-order cumulant of $v_{i\ell}$ is not identified, unless we assume that $v_{i\ell}$ is symmetric (in which case $\widehat{\kappa}_3\left(v_{i\ell}\right)=0$). The fourth-order cumulant of $v_{i\ell}$ can be estimated by

$$\widehat{\kappa}_4\left(v_{i\ell}\right)=\frac{1}{2}\widehat{\kappa}_4\left(y_{i3}-y_{i2}\right), \tag{A5}$$

where the right-hand side in (A5) is simply an empirical fourth-order cumulant. Using (A5) and the symmetry assumption, one can estimate the cumulants of α_i and β_i.

In this example, it is possible to compute simple estimates of the cumulants of β_i that do not require the symmetry assumption. Indeed, taking first differences we get:

$$\begin{aligned}y_{i1}-y_{i2} &= \beta_i+v_{i1}-v_{i2},\\ y_{i2}-y_{i3} &= v_{i2}-v_{i3}.\end{aligned}$$

This motivates the estimators:

$$\widehat{\kappa}_3\left(\beta_i\right) = \widehat{\kappa}_3\left(y_{i1}-y_{i2}\right)-\widehat{\kappa}_3\left(y_{i2}-y_{i3}\right), \tag{A6}$$

$$\widehat{\kappa}_4\left(\beta_i\right) = \widehat{\kappa}_4\left(y_{i1}-y_{i2}\right)-\widehat{\kappa}_4\left(y_{i2}-y_{i3}\right). \tag{A7}$$

B Computing Chamberlain's semiparametric bound

Model and notation. Consider the general panel model that is linear in fixed effects but nonlinear in variables and common parameters:

$$\mathbf{y}_i=\mathbf{a}\left(\mathbf{W}_i,\boldsymbol{\theta}\right)+\mathbf{B}\left(\mathbf{W}_i,\boldsymbol{\theta}\right)\boldsymbol{\gamma}+\mathbf{B}\left(\mathbf{W}_i,\boldsymbol{\theta}\right)\boldsymbol{\varepsilon}_i+\mathbf{v}_i$$

$$\mathbf{E}(\mathbf{v}_i \mid \mathbf{W}_i, \boldsymbol{\gamma}_i) = \mathbf{0}, \qquad \mathbf{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0},$$

where $\boldsymbol{\gamma} = \mathbf{E}(\boldsymbol{\gamma}_i)$ and $\boldsymbol{\varepsilon}_i = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}$. For shortness, write $\mathbf{B}_i = \mathbf{B}(\mathbf{W}_i, \boldsymbol{\theta})$ and $\mathbf{a}_i = \mathbf{a}(\mathbf{W}_i, \boldsymbol{\theta})$. Moreover, let $\mathbf{Var}(\mathbf{y}_i \mid \mathbf{W}_i) = \mathbf{V}_i$, $\mathbf{Var}(\mathbf{v}_i \mid \mathbf{W}_i) = \boldsymbol{\Omega}_i$, and $\mathbf{Var}(\boldsymbol{\varepsilon}_i \mid \mathbf{W}_i) = \boldsymbol{\Sigma}_i$. Thus,

$$\mathbf{V}_i = \mathbf{B}_i \boldsymbol{\Sigma}_i \mathbf{B}_i' + \boldsymbol{\Omega}_i$$

The interest is in the optimal estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ following Chamberlain (1992). For notational simplicity we assume that $\mathbb{S}$ is the full population of individuals.

Optimal estimation of common parameters. Define the idempotent matrix

$$\mathbf{Q}_i = \mathbf{I}_T - \mathbf{B}_i \left(\mathbf{B}_i'\mathbf{B}_i\right)^{-1} \mathbf{B}_i'$$

and let $\mathbf{A}_i$ be a $(T-q) \times T$ semi-triangular matrix such that $\mathbf{Q}_i = \mathbf{A}_i'\mathbf{A}_i$ and $\mathbf{A}_i\mathbf{A}_i' = \mathbf{I}_{T-q}$.

All information about $\boldsymbol{\theta}$ is contained in the $(T-q)$ conditional moments

$$\mathbf{E}(\mathbf{A}_i(\mathbf{y}_i - \mathbf{a}_i) \mid \mathbf{W}_i) = \mathbf{0}.$$

The conditional variance matrix of the transformed residuals is

$$\mathbf{E}\left[\mathbf{A}_i(\mathbf{y}_i - \mathbf{a}_i)(\mathbf{y}_i - \mathbf{a}_i)' \mathbf{A}_i' \mid \mathbf{W}_i\right] = \mathbf{A}_i \boldsymbol{\Omega}_i \mathbf{A}_i' = \mathbf{A}_i \mathbf{V}_i \mathbf{A}_i'.$$

The corresponding optimal instruments are

$$\mathbf{E}\left[\mathbf{D}_i' \left(\mathbf{A}_i\mathbf{V}_i\mathbf{A}_i'\right)^{-1} (\mathbf{A}_i\mathbf{y}_i - \mathbf{A}_i\mathbf{a}_i)\right] = \mathbf{0},$$

where

$$\mathbf{D}_i = \mathbf{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}'} \mathbf{A}_i(\mathbf{y}_i - \mathbf{a}_i) \mid \mathbf{W}_i\right].$$

We show below that

$$\mathbf{D}_i = -\mathbf{A}_i \left(\frac{\partial \mathbf{a}_i}{\partial \boldsymbol{\theta}'} + \sum_{j=1}^{q} \frac{\partial \mathbf{b}_{ji}}{\partial \boldsymbol{\theta}'} \mathbf{E}\left(\gamma_{ji} \mid \mathbf{W}_i\right)\right), \tag{B8}$$

where $\mathbf{B}_i = (\mathbf{b}_{1i}, ..., \mathbf{b}_{qi})$, and $\boldsymbol{\gamma}_i = \left(\gamma_{1i}, ..., \gamma_{qi}\right)'$. Therefore, the optimal moment for $\boldsymbol{\theta}$ is

$$\mathbf{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} [\mathbf{a}_i + \mathbf{B}_i\mathbf{E}(\boldsymbol{\gamma}_i \mid \mathbf{W}_i)]' \mathbf{A}_i' \left(\mathbf{A}_i\mathbf{V}_i\mathbf{A}_i'\right)^{-1} (\mathbf{A}_i\mathbf{y}_i - \mathbf{A}_i\mathbf{a}_i)\right] = \mathbf{0}. \tag{B9}$$

Proof of (B8). We need $\partial \mathbf{A}_i / \partial \theta_k$. First note that the partial derivatives of $\mathbf{Q}_i$ are given by

$$\frac{\partial \mathbf{Q}_i}{\partial \theta_k} = -\mathbf{Q}_i \frac{\partial \mathbf{B}_i}{\partial \theta_k} \left(\mathbf{B}_i'\mathbf{B}_i\right)^{-1} \mathbf{B}_i' - \mathbf{B}_i \left(\mathbf{B}_i'\mathbf{B}_i\right)^{-1} \frac{\partial \mathbf{B}_i'}{\partial \theta_k} \mathbf{Q}_i. \tag{B10}$$

To see the connection between $\mathbf{dQ}_i$ and $\mathbf{dA}_i$ note that

$$\mathbf{dQ}_i = \mathbf{A}_i'(\mathbf{dA}_i) + \left(\mathbf{dA}_i'\right)\mathbf{A}_i$$

$$(\mathbf{dA}_i)\mathbf{A}_i' + \mathbf{A}_i\left(\mathbf{dA}_i'\right) = \mathbf{0},$$

so that

$$\mathbf{A}_i\mathbf{dQ}_i = (\mathbf{dA}_i) + \mathbf{A}_i\left(\mathbf{dA}_i'\right)\mathbf{A}_i = (\mathbf{dA}_i) - (\mathbf{dA}_i)\mathbf{A}_i'\mathbf{A}_i = (\mathbf{dA}_i)\mathbf{B}_i\left(\mathbf{B}_i'\mathbf{B}_i\right)^{-1}\mathbf{B}_i'.$$

Post-multiplying by $\mathbf{B}_i$, the partial derivatives satisfy

$$\mathbf{A}_i \frac{\partial \mathbf{Q}_i}{\partial \theta_k} \mathbf{B}_i = \frac{\partial \mathbf{A}_i}{\partial \theta_k} \mathbf{B}_i.$$

Finally, inserting (B10) and noting that $\mathbf{A}_i \mathbf{B}_i = \mathbf{0}$ it turns out that

$$\frac{\partial \mathbf{A}_i}{\partial \theta_k} \mathbf{B}_i = -\mathbf{A}_i \frac{\partial \mathbf{B}_i}{\partial \theta_k}. \tag{B11}$$

Now, to see that (B8) holds note that

$$\mathbf{D}_i = \mathbf{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}'} \mathbf{A}_i \left(\mathbf{y}_i - \mathbf{a}_i\right) \mid \mathbf{W}_i\right] = \mathbf{E}\left[\begin{array}{ccc} \frac{\partial}{\partial \theta_1} \mathbf{A}_i \left(\mathbf{y}_i - \mathbf{a}_i\right) & \cdots & \frac{\partial}{\partial \theta_K} \mathbf{A}_i \left(\mathbf{y}_i - \mathbf{a}_i\right), \end{array} \mid \mathbf{W}_i\right]$$

and using (B11) we obtain the k-th column of $\mathbf{D}_i$ as follows

$$\begin{aligned}
\mathbf{E}\left[\frac{\partial}{\partial \theta_k} \mathbf{A}_i \left(\mathbf{y}_i - \mathbf{a}_i\right) \mid \mathbf{W}_i\right] &= \left(\frac{\partial \mathbf{A}_i}{\partial \theta_k}\right) \mathbf{E}\left(\mathbf{y}_i - \mathbf{a}_i \mid \mathbf{W}_i\right) - \mathbf{A}_i \left(\frac{\partial \mathbf{a}_i}{\partial \theta_k}\right) \\
&= \left(\frac{\partial \mathbf{A}_i}{\partial \theta_k}\right) \mathbf{E}\left(\mathbf{B}_i \gamma_i + \mathbf{v}_i \mid \mathbf{W}_i\right) - \mathbf{A}_i \left(\frac{\partial \mathbf{a}_i}{\partial \theta_k}\right) \\
&= \left(\frac{\partial \mathbf{A}_i}{\partial \theta_k}\right) \mathbf{B}_i \mathbf{E}\left(\gamma_i \mid \mathbf{W}_i\right) - \mathbf{A}_i \left(\frac{\partial \mathbf{a}_i}{\partial \theta_k}\right) \\
&= -\mathbf{A}_i \left(\frac{\partial \mathbf{a}_i}{\partial \theta_k} + \frac{\partial \mathbf{B}_i}{\partial \theta_k} \mathbf{E}\left(\gamma_i \mid \mathbf{W}_i\right)\right) \\
&= -\mathbf{A}_i \left(\frac{\partial \mathbf{a}_i}{\partial \theta_k} + \sum_{j=1}^{q} \frac{\partial \mathbf{b}_{ji}}{\partial \theta_k} \mathbf{E}\left(\gamma_{ji} \mid \mathbf{W}_i\right)\right).
\end{aligned}$$

Optimal estimation of expected fixed effects. Using matrix inversion formulas, we obtain the following expressions linking $\mathbf{V}_i^{-1}$ and $\boldsymbol{\Omega}_i^{-1}$, which will be used below:

$$\boldsymbol{\Omega}_i^{-1} = \mathbf{V}_i^{-1} + \mathbf{V}_i^{-1} \mathbf{B}_i \left(\boldsymbol{\Sigma}_i^{-1} - \mathbf{B}_i' \mathbf{V}_i^{-1} \mathbf{B}_i\right)^{-1} \mathbf{B}_i' \mathbf{V}_i^{-1}$$

$$\left(\mathbf{B}_i' \mathbf{V}_i^{-1} \mathbf{B}_i\right)^{-1} = \boldsymbol{\Sigma}_i + \left(\mathbf{B}_i' \boldsymbol{\Omega}_i^{-1} \mathbf{B}_i\right)^{-1}.$$

Suppose for the sake of the argument that $\boldsymbol{\theta}$ is known so that $\mathbf{w}_i = \mathbf{y}_i - \mathbf{a}_i$ and $\mathbf{B}_i$ are observable. The model implies the following moments:

$$\mathbf{E}\left[\left(\mathbf{C}_i' \mathbf{B}_i\right)^{-1} \mathbf{C}_i' \left(\mathbf{w}_i - \mathbf{B}_i \gamma\right)\right] = \mathbf{0},$$

for some $\mathbf{C}_i$. So we consider the asymptotic distribution of estimators of the form

$$\widehat{\gamma} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{C}_i' \mathbf{B}_i\right)^{-1} \mathbf{C}_i' \mathbf{w}_i.$$

The scaled estimation error satisfies

$$\sqrt{N}\left(\widehat{\gamma} - \gamma\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{C}_i' \mathbf{B}_i\right)^{-1} \mathbf{C}_i' \mathbf{v}_i \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Upsilon}\right),$$

where

$$\boldsymbol{\Upsilon} = \mathbf{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i'\right) + \mathbf{E}\left[\left(\mathbf{C}_i' \mathbf{B}_i\right)^{-1} \mathbf{C}_i' \boldsymbol{\Omega}_i \mathbf{C}_i \left(\mathbf{B}_i' \mathbf{C}_i\right)^{-1}\right].$$

An optimal choice of $\mathbf{C}_i$ satisfies

$$\left(\mathbf{C}_i'\mathbf{B}_i\right)^{-1}\mathbf{C}_i' = \left(\mathbf{B}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\boldsymbol{\Omega}_i^{-1},$$

which leads to

$$\boldsymbol{\Upsilon} = \mathbf{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\right) + \mathbf{E}\left[\left(\mathbf{B}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{B}_i\right)^{-1}\right], \tag{B12}$$

or

$$\boldsymbol{\Upsilon} = \mathbf{Var}\left[\mathbf{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{W}_i\right)\right] + \mathbf{E}\left[\left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)^{-1}\right]. \tag{B13}$$

One optimal choice is $\mathbf{C}_i' = \mathbf{B}_i'\boldsymbol{\Omega}_i^{-1}$. To characterize the range of optimal choices, let us define $\boldsymbol{\Psi}_i$ for some $q \times q$ matrix $\mathbf{K}_i \geq \mathbf{0}$ such that:

$$\boldsymbol{\Psi}_i^{-1} = \mathbf{V}_i^{-1} + \mathbf{V}_i^{-1}\mathbf{B}_i\mathbf{K}_i\left[\mathbf{I} - \left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)\mathbf{K}_i\right]^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1}$$

Note that setting $\mathbf{K}_i = \boldsymbol{\Sigma}_i$ we have $\boldsymbol{\Psi}_i = \boldsymbol{\Omega}_i$. However, while $\boldsymbol{\Psi}_i$ depends on $\mathbf{K}_i$ the quantity $\left(\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}$ does not:[2]

$$\left(\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1} = \left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1} = \left(\mathbf{B}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\boldsymbol{\Omega}_i^{-1}$$

The conclusion is that an optimal moment uses $\mathbf{C}_i' = \mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}$, and all optimal instruments of the form $\left(\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}$ are the same regardless of the value of $\mathbf{K}_i$. Thus, we can set $\mathbf{K}_i = \mathbf{0}$ without lack of generality and use $\mathbf{C}_i' = \mathbf{B}_i'\mathbf{V}_i^{-1}$.

Therefore, the form of an estimator that attains the bound is

$$\widehat{\gamma} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}\mathbf{w}_i,$$

which is numerically identical for all permissible values of $\mathbf{K}_i$.

The optimal moment conditions for $\boldsymbol{\gamma}$ can be written as

$$\mathbf{E}\left[\left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1}\left(\mathbf{w}_i - \mathbf{B}_i\boldsymbol{\gamma}\right)\right] = \mathbf{0}. \tag{B14}$$

Joint optimal moments: system GMM. It is easy to see that the optimal moments for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, (B9) and (B14) respectively, are uncorrelated:

$$\mathbf{E}\left(\left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1}\left(\mathbf{B}_i\boldsymbol{\varepsilon}_i + \mathbf{v}_i\right)\mathbf{v}_i'\mathbf{A}_i'\left(\mathbf{A}_i\mathbf{V}_i\mathbf{A}_i'\right)^{-1}\mathbf{A}_i\frac{\partial}{\partial\boldsymbol{\theta}'}\left[\mathbf{a}_i + \mathbf{B}_i\mathbf{E}\left(\boldsymbol{\gamma}_i \mid \mathbf{W}_i\right)\right]\right) = \mathbf{0}.$$

Therefore, the optimal moments for estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ are:

$$\mathbf{E}\begin{pmatrix} \frac{\partial}{\partial\boldsymbol{\theta}}\left[\mathbf{a}_i + \mathbf{B}_i\mathbf{E}\left(\boldsymbol{\gamma}_i \mid \mathbf{W}_i\right)\right]'\mathbf{A}_i'\left(\mathbf{A}_i\mathbf{V}_i\mathbf{A}_i'\right)^{-1}\left(\mathbf{A}_i\mathbf{y}_i - \mathbf{A}_i\mathbf{a}_i\right) \\ \left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1}\left(\mathbf{y}_i - \mathbf{a}_i - \mathbf{B}_i\boldsymbol{\gamma}\right) \end{pmatrix} = \mathbf{0}.$$

[2] Note that

$$\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1} = \left[\mathbf{I} - \left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)\mathbf{K}_i\right]^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1}$$

$$\mathbf{B}_i'\boldsymbol{\Psi}_i^{-1}\mathbf{B}_i = \left[\mathbf{I} - \left(\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i\right)\mathbf{K}_i\right]^{-1}\mathbf{B}_i'\mathbf{V}_i^{-1}\mathbf{B}_i.$$

C Semiparametric bound for higher-order cumulants

Consider further extending the model to specify the third-order moments of errors as:

$$\mathbf{E}\left(\mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i | \mathbf{W}_i, \boldsymbol{\gamma}_\mathbf{i}\right) = \boldsymbol{\mu}_{3i}\left(\boldsymbol{\phi}_3\right). \tag{C15}$$

We can write third-order moment restrictions as:

$$\begin{aligned}\mathbf{E}\left(\mathbf{y}_i \otimes \mathbf{y}_i \otimes \mathbf{y}_i | \mathbf{W}_i, \boldsymbol{\gamma}_i\right) &= \left(\mathbf{Z}_i\boldsymbol{\delta} \otimes \mathbf{Z}_i\boldsymbol{\delta} \otimes \mathbf{Z}_i\boldsymbol{\delta}\right) + \mathbf{P}_T\left(\boldsymbol{\psi}_i \otimes \mathbf{Z}_i\boldsymbol{\delta}\right) + \boldsymbol{\mu}_{3i} \\ &\quad + \mathbf{P}_T\left[\left(\mathbf{Z}_i\boldsymbol{\delta} \otimes \mathbf{Z}_i\boldsymbol{\delta} + \boldsymbol{\psi}_i\right) \otimes \mathbf{X}_i\right]\boldsymbol{\gamma}_i + \mathbf{P}_T\left[\mathbf{Z}_i\boldsymbol{\delta} \otimes \mathbf{X}_i \otimes \mathbf{X}_i\right]\left(\boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i\right) \\ &\qquad + \left(\mathbf{X}_i \otimes \mathbf{X}_i \otimes \mathbf{X}_i\right)\left(\boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i\right),\end{aligned}$$

where $\boldsymbol{\psi}_i = \boldsymbol{\psi}_i\left(\boldsymbol{\phi}_2\right)$, $\boldsymbol{\mu}_{3i} = \boldsymbol{\mu}_{3i}\left(\boldsymbol{\phi}_3\right)$, and $\mathbf{P}_T$ denotes the $T^3 \times T^3$ "triplicating" permutation matrix that satisfies, for all $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in \mathbf{R}^{3T}$:

$$\mathbf{P}_T\left(\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}\right) = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} + \mathbf{b} \otimes \mathbf{c} \otimes \mathbf{a} + \mathbf{c} \otimes \mathbf{a} \otimes \mathbf{b}.$$

We can then stack first, second, and third-order moment restrictions to obtain:

$$\mathbf{E}\left(\mathbf{y}_i^{3*} | \mathbf{W}_i, \boldsymbol{\gamma}_i^{3*}\right) = \mathbf{d}_3\left(\mathbf{W}_i, \boldsymbol{\theta}_3\right) + \mathbf{R}_3\left(\mathbf{W}_i, \boldsymbol{\theta}_3\right)\boldsymbol{\gamma}_i^{3*}, \tag{C16}$$

where $\boldsymbol{\theta}_3 = \left(\boldsymbol{\delta}, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3\right)$, and:

$$\boldsymbol{\gamma}_i^{3*} = \begin{pmatrix} \boldsymbol{\gamma}_i \\ \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \\ \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \end{pmatrix}.$$

Equation (C16) still falls into the framework considered in Chamberlain (1992). Note that this approach can be extended to the m-th order, yielding:

$$\mathbf{E}\left(\mathbf{y}_i^{m*} | \mathbf{W}_i, \boldsymbol{\gamma}_i^{m*}\right) = \mathbf{d}_m\left(\mathbf{W}_i, \boldsymbol{\theta}_m\right) + \mathbf{R}_m\left(\mathbf{W}_i, \boldsymbol{\theta}_m\right)\boldsymbol{\gamma}_i^{m*}, \tag{C17}$$

where $\boldsymbol{\theta}_m = \left(\boldsymbol{\delta}, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3, ..., \boldsymbol{\phi}_m\right)$, with $\boldsymbol{\phi}_3, ..., \boldsymbol{\phi}_m$ a parameterization of error moments up to the m-th order, and where:

$$\boldsymbol{\gamma}_i^{m*} = \begin{pmatrix} \boldsymbol{\gamma}_i \\ \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \\ \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i \\ ... \\ \underbrace{\boldsymbol{\gamma}_i \otimes ... \otimes \boldsymbol{\gamma}_i}_{\text{m times}} \end{pmatrix}.$$

This framework can be used to compute semiparametric efficiency bounds under the independence assumption between individual effects and errors (Assumption **??**). We focus on computing bounds for $\boldsymbol{\delta}$, although any moment of individual effects or errors could be analyzed in a similar way.

Consider the increasing sequence of moment conditions (C17), for $m = 2, 3, ...$ Let $\mathbf{V}^{(m)}$ be the efficiency bound on the asymptotic variance for $\boldsymbol{\delta}$ obtained from the first m of those moment conditions. $\mathbf{V}^{(m)}$ can be computed using Chamberlain (1992)'s results. Following the discussion in the previous section, $\mathbf{V}^{(m)}$ is the efficiency bound corresponding to the conditional moment restriction:

$$\mathbf{E}\left[\mathbf{A}_{mi}\left(\mathbf{y}_i^{m*} - \mathbf{d}_m\left(\mathbf{W}_i, \boldsymbol{\theta}_m\right)\right) | \mathbf{W}_i\right] = 0,$$

where $\mathbf{A}_{mi}$ is a generalized orthogonal deviation operator such that:

$$\mathbf{A}_{mi}'\mathbf{A}_{mi} = \mathbf{I} - \mathbf{R}_m(\mathbf{W}_i, \boldsymbol{\theta}_m)\left(\mathbf{R}_m(\mathbf{W}_i, \boldsymbol{\theta}_m)'\mathbf{R}_m(\mathbf{W}_i, \boldsymbol{\theta}_m)\right)^{-1}\mathbf{R}_m(\mathbf{W}_i, \boldsymbol{\theta}_m)'.$$

The sequence $\mathbf{V}^{(m)}$ being nonincreasing in the semi-definite sense (as a larger m means that a larger number of moment conditions is used), we can define the limit:[3]

$$\mathbf{V}^{(\infty)} = \lim_{m\to+\infty}\mathbf{V}^{(m)}.$$

Let $\mathbf{V}_0$ be the semiparametric bound for $\boldsymbol{\delta}$ under independence. Clearly, as $\mathbf{V}_0 \leq \mathbf{V}^{(m)}$ for all m, it follows that $\mathbf{V}_0 \leq \mathbf{V}^{(\infty)}$.

Newey (2004) studies conditions under which the asymptotic variance of the optimal GMM estimator based on an increasing sequence of conditional moment conditions tends to the semiparametric bound, that is, when $\mathbf{V}_0 = \mathbf{V}^{(\infty)}$. He finds that for this to hold, a *spanning* condition is sufficient. This condition requires that the restrictions imposed by the moment conditions are equivalent to those imposed by the semiparametric model.

Intuitively, we expect a *spanning* condition to hold in our case, as the increasing sequence of moment conditions (C17) exhausts all the restrictions implied by independence. We therefore conjecture that $\mathbf{V}_0 = \mathbf{V}^{(\infty)}$.

D Consistent standard errors for the linear projection coefficients

In this section of the appendix we assume that $\mathbb{S}$ is the full population, in order to simplify the notation. The regression coefficients in:

$$\gamma_{\ell i} = \mathbf{F}_i'\boldsymbol{\pi}_\ell + \xi_{\ell i}, \quad \ell = 1, ..., q \tag{D18}$$

where $\mathbf{F}_i$ is such that $\mathbf{E}(\mathbf{v}_i|\mathbf{W}_i, \mathbf{F}_i) = \mathbf{0}$, are given by

$$\boldsymbol{\pi}_\ell = \left[\mathbf{E}\left(\mathbf{F}_i\mathbf{F}_i'\right)\right]^{-1}\mathbf{E}\left(\mathbf{F}_i\gamma_{\ell i}\right), \tag{D19}$$

and a root-N-consistent estimator of $\boldsymbol{\pi}_\ell$ is

$$\widehat{\boldsymbol{\pi}}_\ell = \left(\frac{1}{N}\sum_{i=1}^N \mathbf{F}_i\mathbf{F}_i'\right)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{F}_i\widetilde{\gamma}_{\ell i}, \tag{D20}$$

where, if $\mathbf{h}_{i\ell}'$ denotes the ℓth row of matrix $\mathbf{H}_i$:

$$\widetilde{\gamma}_{\ell i} \equiv \mathbf{h}_{i\ell}'\left(\mathbf{y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\delta}}\right).$$

We have:

$$\begin{aligned}\widetilde{\gamma}_{\ell i} &= \mathbf{h}_{i\ell}'\left(\mathbf{Z}_i\boldsymbol{\delta} + \mathbf{X}_i\boldsymbol{\gamma}_i + \mathbf{v}_i - \mathbf{Z}_i\widehat{\boldsymbol{\delta}}\right)\\ &= \mathbf{F}_i'\boldsymbol{\pi}_\ell + \xi_{\ell i} - \mathbf{h}_{i\ell}'\mathbf{Z}_i\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) + \mathbf{h}_{i\ell}'\mathbf{v}_i.\end{aligned}$$

[3] See Lemma B.1 in Newey (2004).

Hence, letting $\mathbf{\Psi}_N = N^{-1}\sum_{i=1}^N \mathbf{F}_i\mathbf{F}_i'$ we have

$$\mathbf{\Psi}_N\left(\widehat{\boldsymbol{\pi}}_\ell - \boldsymbol{\pi}_\ell\right) = \left(\frac{1}{N}\sum_{i=1}^N \mathbf{F}_i\xi_{\ell i}\right) - \left(\frac{1}{N}\sum_{i=1}^N \mathbf{F}_i\mathbf{h}_{i\ell}'\mathbf{Z}_i\right)\left(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) + \left(\frac{1}{N}\sum_{i=1}^N \mathbf{F}_i\mathbf{h}_{i\ell}'\mathbf{v}_i\right).$$

Also

$$\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta} = \left(\frac{1}{N}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{Q}_i\mathbf{Z}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{Q}_i\mathbf{v}_i. \tag{D21}$$

It is easily shown (e.g., Wooldridge, 2002, p.321 for a special case) that a consistent estimator of $\mathbf{Avar}\left[\sqrt{N}\left(\widehat{\boldsymbol{\pi}}_\ell - \boldsymbol{\pi}_\ell\right)\right]$ is given by:

$$\mathbf{\Psi}_N^{-1}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{a}_i\mathbf{a}_i'\right)\mathbf{\Psi}_N^{-1},$$

where

$$\mathbf{a}_i = \mathbf{F}_i\left(\mathbf{h}_{i\ell}'\left(\mathbf{y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\delta}}\right) - \mathbf{F}_i'\widehat{\boldsymbol{\pi}}_\ell\right) - \left(\sum_{j=1}^N \mathbf{F}_j\mathbf{h}_{j\ell}'\mathbf{Z}_j\right)\left(\sum_{j=1}^N \mathbf{Z}_j'\mathbf{Q}_j\mathbf{Z}_j\right)^{-1}\mathbf{Z}_i'\mathbf{Q}_i\left(\mathbf{y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\delta}}\right).$$

E Examples

In this section of the appendix we provide additional details about the two examples that we use as illustration in the text.

Example 1. We have:

$$\begin{aligned}\underbrace{y_{it} - \rho y_{i,t-1}}_{y_{it}^*(\rho)} &= (1-\rho)\alpha_i + \beta_i\left(t - \rho(t-1)\right) + u_{it} \\ &= \underbrace{(1-\rho)\alpha_i + \rho\beta_i}_{\alpha_i^*(\rho)} + \underbrace{t(1-\rho)\beta_i}_{t\beta_i^*(\rho)} + u_{it}.\end{aligned}$$

When $T = 4$, we obtain the following covariance restrictions:[4]

$$\begin{cases} \text{Var}\left(y_{i2}^*(\rho)\right) = \text{Var}\left(\alpha_i^*(\rho)\right) + 4\,\text{Cov}\left(\alpha_i^*(\rho), \beta_i^*(\rho)\right) \\ \qquad +4\,\text{Var}\left(\beta_i^*(\rho)\right) + \text{Var}\left(u_{i2}\right), \\ \text{Var}\left(y_{i3}^*(\rho) - y_{i2}^*(\rho)\right) = \text{Var}\left(\beta_i^*(\rho)\right) + \text{Var}\left(u_{i3} - u_{i2}\right), \\ \text{Var}\left(y_{i4}^*(\rho) - 2y_{i3}^*(\rho) + y_{i2}^*(\rho)\right) = \text{Var}\left(u_{i4} - 2u_{i3} + u_{i2}\right), \\ \text{Cov}\left(y_{i2}^*(\rho), y_{i3}^*(\rho) - y_{i2}^*(\rho)\right) = \text{Cov}\left(\alpha_i^*(\rho), \beta_i^*(\rho)\right) + 2\,\text{Var}\left(\beta_i^*(\rho)\right) \\ \qquad + \text{Cov}\left(u_{i2}, u_{i3} - u_{i2}\right), \\ \text{Cov}\left(y_{i2}^*(\rho), y_{i4}^*(\rho) - 2y_{i3}^*(\rho) + y_{i2}^*(\rho)\right) = \text{Cov}\left(u_{i2}, u_{i4} - 2u_{i3} + u_{i2}\right), \\ \text{Cov}\left(y_{i3}^*(\rho) - y_{i2}^*(\rho), y_{i4}^*(\rho) - 2y_{i3}^*(\rho) + y_{i2}^*(\rho)\right) = \text{Cov}\left(u_{i3} - u_{i2}, u_{i4} - 2u_{i3} + u_{i2}\right). \end{cases}$$

Note that ρ is not identified from levels equations when $T = 4$. When $T = 5$ we obtain additional identifying restrictions which may suffice for ρ to be identified. For example:

$$\begin{aligned}\text{Var}\left(\beta_i^*(\rho)\right) &= \text{Cov}\left[y_{i3}^*(\rho) - y_{i2}^*(\rho), y_{i5}^*(\rho) - y_{i4}^*(\rho)\right] \\ &= \text{Var}\left(y_{i3}^*(\rho) - y_{i2}^*(\rho)\right) + \frac{1}{2}\text{Cov}\left[y_{i3}^*(\rho) - 2y_{i2}^*(\rho), y_{i4}^*(\rho) - 2y_{i3}^*(\rho) + y_{i2}^*(\rho)\right].\end{aligned}$$

[4] $T = 4$ means that we have 3 observations on $y_{it}^*(\rho)$ for given ρ $(t = 2, 3, 4)$.

Example 2. Covariance restrictions in levels are:

$$\begin{cases} \text{Var}(y_{i1}) & = & \text{Var}(\alpha_i) + 2\,\text{Cov}(\alpha_i, \beta_i) + \text{Var}(\beta_i) + \text{Var}(v_{i1}), \\ \text{Var}(y_{i2}) & = & \text{Var}(\alpha_i) + \text{Var}(v_{i2}), \\ \text{Var}(y_{i3}) & = & \text{Var}(\alpha_i) + \text{Var}(v_{i3}), \\ \text{Cov}(y_{i1}, y_{i2}) & = & \text{Var}(\alpha_i) + \text{Cov}(\alpha_i, \beta_i) + \text{Cov}(v_{i1}, v_{i2}), \\ \text{Cov}(y_{i1}, y_{i3}) & = & \text{Var}(\alpha_i) + \text{Cov}(\alpha_i, \beta_i) + \text{Cov}(v_{i1}, v_{i3}), \\ \text{Cov}(y_{i2}, y_{i3}) & = & \text{Var}(\alpha_i) + \text{Cov}(v_{i2}, v_{i3}). \end{cases}$$

F Testing the covariance structure of errors

In practice, it may be important to empirically determine the order of the MA process of the error terms. This is of special importance in order to estimate the variance of individual effects, as misspecifying the form of the variance matrix of errors would result in inconsistent estimates. This can be done easily using the results of the paper, as we now explain.

Let $\mathbf{S}_2$ be a selection matrix with m columns, and suppose that one wants to test

$$H_0 : \mathbf{vec}(\boldsymbol{\Omega}_i) = \mathbf{S}_2 \boldsymbol{\omega}_i$$

against an unrestricted alternative. We have, under H_0:

$$\mathbf{M}_i \mathbf{E}\left[(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}) \otimes (\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}) | \mathbf{W}_i\right] = \mathbf{M}_i \mathbf{S}_2 \left(\mathbf{M}_i \mathbf{S}_2\right)^{\dagger} \mathbf{M}_i \mathbf{E}\left[(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}) \otimes (\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}) | \mathbf{W}_i\right]. \quad \text{(F22)}$$

This suggests to consider a test of significance of the following quantity:

$$\widehat{\mathcal{T}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{G}_i \mathbf{M}_i \left(\boldsymbol{I}_{T^2} - \mathbf{S}_2 \left(\mathbf{M}_i \mathbf{S}_2\right)^{\dagger} \mathbf{M}_i \right) \left(\widehat{\mathbf{v}}_i \otimes \widehat{\mathbf{v}}_i\right),$$

where $\mathbf{G}_i$ is a $\left(\frac{T(T+1)}{2} - \frac{q(q+1)}{2}\right) \times T^2$ matrix such that $\mathbf{M}_i \mathbf{D}_T = \mathbf{G}_i' \mathbf{C}_i$, with $\mathbf{D}_T$ the duplication matrix (Magnus and Neudecker, 1988, p.49), and $\mathbf{C}_i$ a full row matrix.[5]

The minimum chi-square statistic then satisfies:

$$\widehat{\mathcal{T}}' \widehat{\mathcal{V}}^{-1} \widehat{\mathcal{T}} \xrightarrow{d} \chi^2_d,$$

where $d = \frac{T(T+1)}{2} - \frac{q(q+1)}{2} - m$, and where the matrix $\widehat{\mathcal{V}}$ depends on fourth-order moments of the data.

This strategy may be interpreted as an extension of the test of covariance structures proposed in Abowd and Card (1989) to random coefficients models. In particular, it is immediate to extend the approach to sequentially test various MA structures, starting with the less restrictive one (e.g., testing MA(q), then MA(q-1), etc...). However, a distinctive feature of our test relative to Abowd and Card is that it also incorporates information in levels (see the discussion in Arellano, 2003, p.67).

G Covariance restrictions: order conditions

The following result is useful to obtain the order conditions for identification of variances of effects and errors (Section 3 of the paper).

[5] Note that transformation by $\mathbf{G}_i$ eliminates redundancies.

Lemma G1 *Let* $\mathbf{P}$ *be a symmetric idempotent* $n \times n$ *matrix with rank* p*. Let* $\mathbf{D}_n$ *be the* $n^2 \times n(n+1)/2$ duplication *matrix that transforms* $\text{vech}(\mathbf{A})$ *into* $\text{vec}(\mathbf{A})$*, for any* $n \times n$ *matrix* $\mathbf{A}$ *(Magnus and Neudecker, 1988, p.49). Then:*

i) $\text{rank}\left[\left(\mathbf{I}_{n^2} - \mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n\right] = \frac{n(n+1)}{2} - \frac{p(p+1)}{2}$.

ii) $\text{rank}\left\{\left[\left(\mathbf{I}_n - \mathbf{P}\right) \otimes \left(\mathbf{I}_n - \mathbf{P}\right)\right]\mathbf{D}_n\right\} = \frac{(n-p)(n-p+1)}{2}$,

Proof. Part *i*). The proof uses results from Magnus and Neudecker (1988, MN hereafter). From MN's Theorem 13 p.49-50 we have:

$$\begin{aligned}\left(\mathbf{I}_{n^2} - \mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n &= \mathbf{D}_n\mathbf{D}_n^{\dagger}\left(\mathbf{I}_{n^2} - \mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n \\ &= \mathbf{D}_n\left(\mathbf{I}_{\frac{n(n+1)}{2}} - \mathbf{D}_n^{\dagger}\left(\mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n\right),\end{aligned}$$

where $\mathbf{D}_n^{\dagger} = \left(\mathbf{D}_n'\mathbf{D}_n\right)^{-1}\mathbf{D}_n'$ denotes the Moore-Penrose generalized inverse of $\mathbf{D}_n$.

Hence, because $\mathbf{D}_n$ has full column rank, the rank of: $\left(\mathbf{I}_{n^2} - \mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n$ is equal to that of: $\mathbf{B}_n = \mathbf{I}_{\frac{n(n+1)}{2}} - \mathbf{D}_n^{\dagger}\left(\mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n$. But, using equations (14) and (15) in MN (Theorem 13 p.50) it is easy to show that $\mathbf{B}_n$ is idempotent. So, using MN's Theorem 21 (p.20): $\text{rank}(\mathbf{B}_n) = \text{Tr}(\mathbf{B}_n)$. Now:

$$\begin{aligned}\text{Tr}\left(\mathbf{D}_n^{\dagger}\left(\mathbf{P} \otimes \mathbf{P}\right)\mathbf{D}_n\right) &= \text{Tr}\left(\mathbf{D}_n\mathbf{D}_n^{\dagger}\left(\mathbf{P} \otimes \mathbf{P}\right)\right) \\ &= \frac{1}{2}\text{Tr}\left(\mathbf{P} \otimes \mathbf{P}\right) + \frac{1}{2}\text{Tr}\left(\mathbf{K}_n\left(\mathbf{P} \otimes \mathbf{P}\right)\right) \\ &= \frac{p^2}{2} + \frac{1}{2}\text{Tr}\left(\mathbf{K}_n\left(\mathbf{P} \otimes \mathbf{P}\right)\right),\end{aligned}$$

where $\mathbf{K}_n$ is the *commutation* matrix (MN, p.47). Let $\mathbf{E}_{ij}$ be a $n \times n$ matrix with zeros everywhere, except a one at position (i, j). Let also $\mathbf{P} = [p_{ij}]_{(i,j)}$.

$$\begin{aligned}\text{Tr}\left(\mathbf{K}_n\left(\mathbf{P} \otimes \mathbf{P}\right)\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{vec}\left(\mathbf{E}_{ij}\right)'\mathbf{K}_n\left(\mathbf{P} \otimes \mathbf{P}\right)\mathbf{vec}\left(\mathbf{E}_{ij}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{vec}\left(\mathbf{E}_{ij}\right)'\mathbf{vec}\left(\mathbf{P}\mathbf{E}_{ij}'\mathbf{P}'\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}p_{ji} \\ &= \sum_{i=1}^{n}p_{ii} = p,\end{aligned}$$

where the next to last equality comes from idempotence of $\mathbf{P}$. So:

$$\text{Tr}\left(\mathbf{B}_n\right) = \frac{n(n+1)}{2} - \frac{p^2}{2} - \frac{p}{2}.$$

This ends the proof.

Part *ii*). Because of idempotence: $\text{rank}\left(\mathbf{I}_n - \mathbf{P}\right) = n - p$. Let $\mathbf{v}_1, ..., \mathbf{v}_{n-p}$ be a basis of the vector space spanned by the columns of $\mathbf{I}_n - \mathbf{P}$. Clearly, $\{\mathbf{v}_i \otimes \mathbf{v}_j, (i, j) \in \{1, ..., n-p\}^2\}$ forms a linearly independent family. So does $\{\mathbf{v}_i \otimes \mathbf{v}_j, (i, j) \in \{1, ..., n-p\}^2, i \leq j\}$. As this family has $(n-p)(n-p+1)/2$ elements, the conclusion follows.

■

Part i) implies that the order condition for identification of covariances using restrictions in levels is $T(T+1)/2 - q(q+1)/2 \geq m$.

Alternatively, one may work with within-group equations alone, see equation (7) in the paper. This implies the following covariance restrictions:

$$\mathbf{Q}_i \mathbf{E}\left[\left(\mathbf{y}_i - \mathbf{Z}_i \boldsymbol{\delta}\right)\left(\mathbf{y}_i - \mathbf{Z}_i \boldsymbol{\delta}\right)' \left| \mathbf{W}_i \right.\right] \mathbf{Q}_i' = \mathbf{Q}_i \boldsymbol{\Omega}_i \mathbf{Q}_i'. \tag{G23}$$

Matrix $\boldsymbol{\Omega}_i$ is thus identified from the within-group covariance restrictions (G23) alone, provided:

$$\operatorname{rank}\left[\left(\mathbf{Q}_i \otimes \mathbf{Q}_i\right) \mathbf{S}_2\right] = m. \tag{G24}$$

From Part ii) in Lemma G1, the associated order condition is:

$$\frac{(T-q)(T-q+1)}{2} \geq m.$$

Hence, the order condition is more restrictive than the one using covariance restrictions in levels.

For example, consider an AR(1) model with a single heterogeneous intercept, and $T = 3$. The autoregressive parameter ρ is not identified from within-group equations alone. However, ρ is identified from covariance restrictions in levels, as the IV estimand in the regression of $(y_{i3} - y_{i2})$ on $(y_{i2} - y_{i1})$ using y_{i1} as an instrument.

H Multivariate cumulants and characteristic functions

Here we collect some standard definitions and properties of cumulants and characteristic functions that are used in the paper.

Cumulants. Let $\mathbf{U} = (U_1, ..., U_n)'$ be an n-dimensional random vector with zero mean and well-defined moments to the fourth-order. We define its *cumulant vector of order 3* as the n^3-dimensional vector $\boldsymbol{\kappa}_3(\mathbf{U})$ whose elements $\kappa_3^{i,j,k}(\mathbf{U})$, for $(i,j,k) \in \{1,...,n\}^3$, are arranged in lexicographic order and are such that

$$\kappa_3^{i,j,k}(\mathbf{U}) = \mathbf{E}\left(U_i U_j U_k\right), \quad (i,j,k) \in \{1,...,n\}^3.$$

Likewise, we define $\boldsymbol{\kappa}_4(\mathbf{U})$ whose n^4 elements are

$$\begin{aligned} \kappa_4^{i,j,k,\ell}(\mathbf{U}) &= \mathbf{E}\left(U_i U_j U_k U_\ell\right) - \mathbf{E}\left(U_i U_j\right) \mathbf{E}\left(U_k U_\ell\right) \\ &\quad - \mathbf{E}\left(U_i U_k\right) \mathbf{E}\left(U_j U_\ell\right) - \mathbf{E}\left(U_i U_\ell\right) \mathbf{E}\left(U_j U_k\right), \quad (i,j,k,\ell) \in \{1,...,n\}^4. \end{aligned}$$

For a nonzero mean random vector $\mathbf{V}$, we define $\boldsymbol{\kappa}_3(\mathbf{V}) = \boldsymbol{\kappa}_3\left(\mathbf{V} - \mathbf{E}\left(\mathbf{V}\right)\right)$, and we similarly define $\boldsymbol{\kappa}_4(\mathbf{V})$.

The *skewness* of U_j ($i \in \{1,...,n\}$) and its *kurtosis* are given by: $\kappa_3^{j,j,j}(\mathbf{U}) / \operatorname{Var}(U_j)^{3/2}$ and $\left[\kappa_4^{j,j,j,j}(\mathbf{U}) / \operatorname{Var}(U_j)^2\right] + 3$, respectively. We may similarly define conditional cumulants by replacing the expectations in these formulas by conditional expectations.

Cumulants satisfy a multilinearity property, and can be interpreted as tensors (Kofidis and Regalia, 2000). Namely, for any $s \times n$ matrix $\mathbf{A}$ we have:

$$\begin{aligned} \boldsymbol{\kappa}_3(\mathbf{AU}) &= \left(\mathbf{A} \otimes \mathbf{A} \otimes \mathbf{A}\right) \boldsymbol{\kappa}_3(\mathbf{U}), \\ \boldsymbol{\kappa}_4(\mathbf{AU}) &= \left(\mathbf{A} \otimes \mathbf{A} \otimes \mathbf{A} \otimes \mathbf{A}\right) \boldsymbol{\kappa}_4(\mathbf{U}). \end{aligned}$$

Moreover, cumulants of the sums of *independent* random variables satisfy: $\boldsymbol{\kappa}_3(\mathbf{U} + \mathbf{V}) = \boldsymbol{\kappa}_3(\mathbf{U}) + \boldsymbol{\kappa}_3(\mathbf{V})$, and: $\boldsymbol{\kappa}_4(\mathbf{U} + \mathbf{V}) = \boldsymbol{\kappa}_4(\mathbf{U}) + \boldsymbol{\kappa}_4(\mathbf{V})$. Because of these properties, it is sometimes more convenient to work with cumulants than with moments, although there exists a mapping between the two.

Here we have only defined cumulants of order 3 and 4. We could easily generalize these results to cumulants of order 5 or higher. The first-order cumulant is simply the mean, and the cumulants of order 2 are the variances and covariances.

Characteristic functions. Let $(\mathbf{Y}, \mathbf{X})$ be a pair of random vectors, $\mathbf{Y} \in \mathbf{R}^L$, and let j be a square root of -1. The conditional characteristic function of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$, is defined as:

$$\Psi_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{\tau}|\mathbf{x}) = \mathbf{E}\left(\exp(j\boldsymbol{\tau}'\mathbf{Y})|\mathbf{X} = \mathbf{x}\right), \quad \boldsymbol{\tau} \in \mathbf{R}^L.$$

We make use of the following properties of characteristic functions in the paper (e.g., Dudley, 2002, Chapter 9). First, characteristic functions uniquely determine distribution functions. Moreover, when $\mathbf{Y}|\mathbf{X}$ admits an absolutely continuous density, the mapping between the (conditional) characteristic function and the (conditional) density is given by the *inverse Fourier transform*:

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^L} \int \exp\left(-j\boldsymbol{\tau}'\mathbf{y}\right) \Psi_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{\tau}|\mathbf{x})\mathbf{d}\boldsymbol{\tau}. \tag{H25}$$

Second, if $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent given $\mathbf{X}$ then:

$$\Psi_{\mathbf{Y}_1+\mathbf{Y}_2|\mathbf{X}}(\boldsymbol{\tau}|\mathbf{x}) = \Psi_{\mathbf{Y}_1|\mathbf{X}}(\boldsymbol{\tau}|\mathbf{x})\Psi_{\mathbf{Y}_2|\mathbf{X}}(\boldsymbol{\tau}|\mathbf{x}). \tag{H26}$$

Lastly, cumulants (when they exist) can be obtained from the successive derivatives of the logarithm of the characteristic function (also called cumulant generating function) evaluated at $\boldsymbol{\tau} = \mathbf{0}$.

References

[1] Abowd, J., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica,* 57, 411-445.

[2] Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press.

[3] Chamberlain, G. (1992): "Efficiency Bounds for Semiparametric Regression", *Econometrica*, 60, 567–596.

[4] Dudley (2002): *Real Analysis and Probability*, Cambridge University Press.

[5] Kofidis, E., and P.A. Regalia (2000): "Tensor Approximation and Signal Processing Applications," in: *AMS Conf. on Structured Matrices in Operator Theory, Numerical Analysis, Control, Signal and Image Processing*, AMS Publ.

[6] Magnus, J.R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.

[7] Newey, W. (2004): "Efficient Semiparametric Estimation via Moment Restrictions," *Econometrica*, 72(6), 1877-1897.

[8] Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.